NEAR-OPTIMAL MEAN VALUE ESTIMATES FOR MULTIDIMENSIONAL WEYL SUMS

S. T. PARSELL, S. M. PRENDIVILLE, AND T. D. WOOLEY*

Abstract. We obtain sharp estimates for multidimensional generalisations of Vinogradov's mean value theorem for arbitrary translation-dilation invariant systems, achieving constraints on the number of variables approaching those conjectured to be the best possible. Several applications of our bounds are discussed.

1. Introduction

The investigation of Diophantine problems of large degree is in general fraught with difficulties only partially mollified by the presence of intrinsic diagonal structure. Indeed, such analyses as are made available via the Hardy-Littlewood (circle) method, when successful, involve complicated exponential sum estimates widely considered to be amongst the most challenging in the subject. The application of Weyl differencing by Davenport, Birch and Schmidt, on the one hand, involves a delicate interplay between the singular locus associated with the problem and the quality of ensuing exponential sum estimates (see [4], [5], [13]). As an inescapable feature of such approaches, the number of variables required in a successful treatment grows exponentially with the degree of the problem at hand. The extension of Vinogradov's methods to exponential sums in many variables, on the other hand, is notoriously complicated. The work of Arkhipov, Karatsuba and Chubarikov [1], [2], for example, permits substantially sharper conclusions to be drawn when partial diagonal structure is present. However, the complexity of the underlying methods has deterred a consideration of all but the simplest model situations (see [2] and [9]). In addition, the available conclusions fail to achieve their conjectured potential by a factor growing roughly like the logarithm of the total degree of the associated translation-invariant Diophantine system.

Our goal in this paper is to extend the efficient congruencing method introduced by the third author [21] so as to accommodate the generalised Vinogradov systems of Arkhipov, Karatsuba and Chubarikov (see [1], [2]). It transpires that for systems of large degree, the bounds that we thereby derive miss those conjectured to hold by a factor of only 2 or thereabouts, transforming

2010 *Mathematics Subject Classification.* 11L15, 11L07, 11D45, 11D72, 11P55.

Key words and phrases. Exponential sums, Hardy-Littlewood method, Diophantine equations.

STP was supported by National Security Agency Grant H98230-11-1-0190, SMP by an EPSRC doctoral training grant through the University of Bristol, and TDW by a Royal Society Wolfson Research Merit Award.

the previous state of the art. Moreover, our methods are of such flexibility that they may be successfully applied to translation-invariant systems of wide generality, and in particular to systems closely related to those subject to recent investigations by quantitative arithmetic geometers studying the Manin-Peyre conjectures (see [14, §4.15], [15]). Since our methods yield estimates no less striking for such systems, we take the opportunity to derive rather general estimates again coming within a constant factor of those conjectured to hold. There are consequences of all of this work for exponential sum estimates of Weyl-type, for the solubility of systems of Diophantine equations and related problems, and for certain problems in additive combinatorics, and these we also explore herein.

Rather than encumber the reader at this point with the substantial notational prerequisites entailed by a discussion of our most general conclusions, we instead offer the more easily digestible corollaries particular to the model problem considered in earlier work [9] of the first author. Let s, k and d be natural numbers, and let X be a positive number. We focus attention on the system of simultaneous Diophantine equations

$$\sum_{j=1}^{s} x_{j1}^{i_1} x_{j2}^{i_2} \cdots x_{jd}^{i_d} = \sum_{j=1}^{s} y_{j1}^{i_1} y_{j2}^{i_2} \cdots y_{jd}^{i_d} \quad (1 \leqslant i_1 + \ldots + i_d \leqslant k). \tag{1.1}$$

Here, the indices i_m are non-negative integers, so that a modest computation reveals the total number of equations in the system (1.1) to be $r = r_{d,k}$, where

$$r_{d,k} = \binom{k+d}{d} - 1. \tag{1.2}$$

Meanwhile, the total degree of the system (1.1), which is to say the sum of the degrees of the equations comprising the system, is equal to $K = K_{d,k}$, where

$$K_{d,k} = \sum_{l=1}^{k} l \binom{l+d-1}{l} = \frac{d}{d+1}(r+1)k. \tag{1.3}$$

In particular, in the familiar classical Vinogradov system with $d = 1$, one has $r = k$ and $K = \frac{1}{2}k(k+1)$. Finally, we write $J_{s,k,d}(X)$ for the number of integral solutions of the system (1.1) with $1 \leqslant x_{jm}, y_{jm} \leqslant X$ $(1 \leqslant j \leqslant s,\ 1 \leqslant m \leqslant d)$.

In §2, as a special case of Theorem 2.1, we derive an estimate for $J_{s,k,d}(X)$ that is in many respects close to the best possible. Here and throughout, implicit constants in Vinogradov's notation $\ll$ and $\gg$ depend at most on s, k, d and ε, unless otherwise indicated. The letter X should be interpreted as a positive number sufficiently large in terms of s, k, d and ε.

Theorem 1.1. *Suppose that s, k and d are natural numbers with $k \geqslant 2$ and $s \geqslant r(k+1)$. Then for each $\varepsilon > 0$, one has $J_{s,k,d}(X) \ll X^{2sd-K+\varepsilon}$.*

The special case of Theorem 1.1 with $d = 1$ is equivalent to [21, Theorem 1.1], a conclusion which has very recently been sharpened in [22, Theorem 1.1], so that for $k \geqslant 3$ and $s \geqslant k^2 - 1$, one has

$$J_{s,k,1}(X) \ll X^{2s-\frac{1}{2}k(k+1)+\varepsilon}.$$

When $d > 1$, meanwhile, one may compare the conclusion of Theorem 1.1 with work of Arkhipov, Karatsuba and Chubarikov [2]. Although set up slightly differently, it is clear that the methods of the latter authors have the potential to establish a bound of the shape

$$J_{s,k,d}(X) \ll X^{2sd-K+\Delta_s},$$

where Δ_s decays with s roughly like $rke^{-s/(rk)}$. In [9, Theorem 1.1] this conclusion was improved by the first author for sufficiently large values of k, so that on writing

$$s_0 = \tfrac{1}{2}rk(\log k - 2\log\log k),$$

one may take

$$\Delta_s = \begin{cases} rke^{2-2s/(rk)}, & \text{for } 1 \leqslant s \leqslant s_0, \\ r(\log k)^2 e^{3-3(s-s_0)/(2rk)}, & \text{for } s > s_0. \end{cases}$$

The estimate supplied by Theorem 1.1 is substantially sharper. Thus, provided only that $s \geqslant r(k+1)$, one may take $\Delta_s = \varepsilon$ for any positive number ε. The number of variables required in typical applications, as we discuss in due course, is thereby reduced by a factor of order $\log(rk)$.

In order to discern the strength of the estimate supplied by Theorem 1.1, we must consider available lower bounds for the mean value $J_{s,k,d}(X)$, and thereby infer plausible conjectures for corresponding upper bounds. In §3 we establish the lower bound for $J_{s,k,d}(X)$ contained in the following theorem.

Theorem 1.2. *Suppose that s, k and d are natural numbers. Then one has*

$$J_{s,k,d}(X) \gg X^{sd} + \sum_{j=1}^{d} X^{(2s-1)j+d-K_{j,k}}.$$

It seems reasonable to conjecture that whenever $\varepsilon > 0$, one has the allied upper bound

$$J_{s,k,d}(X) \ll X^{\varepsilon}\Big(X^{sd} + \sum_{j=1}^{d} X^{(2s-1)j+d-K_{j,k}}\Big).$$

Thus, when s is sufficiently large in terms of k and d, one expects that

$$J_{s,k,d}(X) \ll X^{2sd-K+\varepsilon}, \tag{1.4}$$

as is confirmed by Theorem 1.1 for $s \geqslant r(k+1)$. We emphasise that here, and throughout the introduction, we abbreviate $r_{d,k}$ to r and $K_{d,k}$ to K. As a consequence of Theorem 1.2, we show in Theorem 3.2 that when δ is a real number with $2d/k < \delta < 1$, and

$$s \leqslant \frac{d}{2d+2}(1-\delta)r(k+1),$$

then there is a positive number $\eta = \eta(d,k)$ with the property that

$$J_{s,k,d}(X) \gg X^{2sd-K+\eta}.$$

When $d \geqslant 2$, therefore, it follows that the conclusion of Theorem 1.1 comes within a factor $2+2/d+O(d/k)$ of the least value of s for which the conjectured upper bound (1.4) might conceivably hold. In such multidimensional Weyl sums, a near-optimal conclusion of this type, merely a constant factor away from the best possible, has hitherto been wholly beyond our grasp.

We next consider upper bounds for exponential sums of Weyl-type, the discussion of which is much facilitated by the introduction of additional notation. It is convenient to abbreviate a monomial of the shape $x_1^{i_1}x_2^{i_2}\cdots x_d^{i_d}$ to $\mathbf{x}^{\mathbf{i}}$, in which $\mathbf{i} = (i_1, i_2, \dots, i_d)$. Likewise, we may write $\mathbf{x}_m^{\mathbf{i}}$ for $x_{m1}^{i_1}x_{m2}^{i_2}\cdots x_{md}^{i_d}$. In such circumstances, we put

$$|\mathbf{i}| = i_1 + \ldots + i_d.$$

Also, in place of the s-tuple $(\mathbf{x}_1, \dots, \mathbf{x}_s)$ we write $\overline{\mathbf{x}}$, and we adopt the convention that $a \leqslant \mathbf{v} \leqslant b$ is to mean that each coordinate v_l of the vector $\mathbf{v}$ satisfies $a \leqslant v_l \leqslant b$. Equipped with these conventions, the Diophantine system (1.1) assumes the compact shape

$$\mathbf{x}_1^{\mathbf{i}} + \ldots + \mathbf{x}_s^{\mathbf{i}} = \mathbf{y}_1^{\mathbf{i}} + \ldots + \mathbf{y}_s^{\mathbf{i}} \quad (1 \leqslant |\mathbf{i}| \leqslant k),$$

and $J_{s,k,d}(X)$ counts the number of integral solutions of this system with $1 \leqslant \overline{\mathbf{x}}, \overline{\mathbf{y}} \leqslant X$.

Define the exponential sum $f(\boldsymbol{\alpha}) = f_{d,k}(\boldsymbol{\alpha}; X)$ by

$$f_{d,k}(\boldsymbol{\alpha}; X) = \sum_{1 \leqslant \mathbf{x} \leqslant X} e(\psi(\mathbf{x}; \boldsymbol{\alpha})),$$

where

$$\psi(\mathbf{x}; \boldsymbol{\alpha}) = \sum_{1 \leqslant |\mathbf{i}| \leqslant k} \alpha_{\mathbf{i}} \mathbf{x}^{\mathbf{i}},$$

and, as usual, we write $e(z)$ for $e^{2\pi i z}$. Here, the subscript d that identifies $\mathbf{x}$ as the d-tuple $(x_1, \dots, x_d)$ may usually be omitted without leading to confusion. As we have already noted, the number of coefficients $\alpha_{\mathbf{i}}$ is r. We adopt the convention that, when $G : [0,1)^n \to \mathbb{C}$ is measurable, then

$$\oint G(\boldsymbol{\beta})\, \mathrm{d}\boldsymbol{\beta} = \int_{[0,1)^n} G(\boldsymbol{\beta})\, \mathrm{d}\boldsymbol{\beta}.$$

It then follows from orthogonality that

$$J_{s,k,d}(X) = \oint |f_{d,k}(\boldsymbol{\alpha}; X)|^{2s}\, \mathrm{d}\boldsymbol{\alpha}. \tag{1.5}$$

Upper bounds for mean values of exponential sums such as $f(\boldsymbol{\alpha})$ may be converted into Weyl-type estimates by means of variants of the large sieve inequality. Before announcing such an estimate, which is a consequence of the more general result recorded in Theorem 10.3, we pause to record a further notational convention. When $\mathbf{a} \in \mathbb{Z}^n$, we write $(q, \mathbf{a})$ for the greatest common divisor $(q, a_1, \dots, a_n)$.

Theorem 1.3. *Suppose that d and k are natural numbers with $k \geqslant 2$. Let σ be any real number with*

$$\sigma^{-1} \geqslant \left(2k\binom{k+d-1}{d} - 2k + 1\right)(d+1).$$

Then whenever $|f_{d,k}(\boldsymbol{\alpha}; X)| \geqslant X^{d-\sigma+\varepsilon}$, for some $\varepsilon > 0$, it follows that there exist $q \in \mathbb{N}$ and $a_{\mathbf{j}} \in \mathbb{Z}$ $(1 \leqslant |\mathbf{j}| \leqslant k)$ satisfying

$$(q, \mathbf{a}) = 1, \quad 1 \leqslant q \leqslant X^{k\sigma} \quad \text{and} \quad |q\alpha_{\mathbf{j}} - a_{\mathbf{j}}| \leqslant X^{k\sigma - |\mathbf{j}|} \quad (1 \leqslant |\mathbf{j}| \leqslant k).$$

The special case of Theorem 1.3 with $d = 1$ is very slightly weaker than [21, Theorem 1.6], a conclusion which has recently been sharpened in [22, Theorem 11.2]. Thus, when $d = 1$, the conclusion of Theorem 1.3 holds whenever $k \geqslant 4$ and $\sigma^{-1} \geqslant 4k(k-2)$. The work of Arkhipov, Karatsuba and Chubarikov [2], as interpreted and sharpened by the first author, yields a conclusion similar to Theorem 1.3. Indeed, it follows from a corrected version[1] of [9, Theorem 1.2] that a conclusion of similar form holds for sufficiently large values of k, though with the constraint on the exponent σ replaced by a condition of the shape $\sigma^{-1} \geqslant \frac{4}{3}(d+1)rk\log(rk)$. On noting that $\binom{k+d-1}{d} - 1 = r_{d,k-1} \leqslant r_{d,k}$, the superiority of our new bound is clear.

We next consider the application of our new estimates to Diophantine problems. When s, k and d are natural numbers, and $a_{\mathbf{i}j}$ is a non-zero integer for $1 \leqslant |\mathbf{i}| \leqslant k$ and $1 \leqslant j \leqslant s$, write

$$\phi_{\mathbf{i}}(\overline{\mathbf{x}}) = \sum_{j=1}^{s} a_{\mathbf{i}j}\mathbf{x}_j^{\mathbf{i}} \quad (1 \leqslant |\mathbf{i}| \leqslant k).$$

In §11, we consider the Diophantine system

$$\phi_{\mathbf{i}}(\overline{\mathbf{x}}) = 0 \quad (1 \leqslant |\mathbf{i}| \leqslant k), \tag{1.6}$$

consisting of r equations of total degree K. Let $N(B)$ denote the number of integral solutions of the system (1.6) with $|\overline{\mathbf{x}}| \leqslant B$. We follow Schmidt [13] when defining the (formal) real and p-adic densities associated with the system (1.6). When $L > 0$, define

$$\lambda_L(\eta) = \begin{cases} L(1 - L|\eta|), & \text{when } |\eta| \leqslant L^{-1}, \\ 0, & \text{otherwise,} \end{cases}$$

and put

$$\mu_L = \int_{|\overline{\boldsymbol{\xi}}| \leqslant 1} \prod_{1 \leqslant |\mathbf{i}| \leqslant k} \lambda_L(\phi_{\mathbf{i}}(\overline{\boldsymbol{\xi}}))\, \mathrm{d}\overline{\boldsymbol{\xi}}.$$

The limit $\sigma_\infty = \lim_{L \to \infty} \mu_L$, when it exists, is called the *real density*. Meanwhile, given a natural number q, we write

$$M(q) = \operatorname{card}\left\{\overline{\mathbf{x}} \in (\mathbb{Z}/q\mathbb{Z})^{sd} : \phi_{\mathbf{i}}(\overline{\mathbf{x}}) \equiv 0 \pmod q \ (1 \leqslant |\mathbf{i}| \leqslant k)\right\}.$$

[1]See the discussion following the proof of Theorem 10.2 below for an explanation of the need for a modest correction in [9, Theorem 1.2].

For each prime number p, we then put

$$\sigma_p = \lim_{h\to\infty} p^{h(r-sd)} M(p^h),$$

provided that this limit exists, and refer to σ_p as the *p-adic density*.

As a special case of Theorem 11.1, we establish an asymptotic formula for $N(B)$ valid whenever $s \geqslant 2r(k+1)+1$.

Theorem 1.4. *Suppose that s, k and d are natural numbers with $k \geqslant 2$ and $s \geqslant 2r(k+1)+1$. In addition, let $a_{\mathbf{i}j}$ $(1 \leqslant |\mathbf{i}| \leqslant k, 1 \leqslant j \leqslant s)$ be non-zero integers. Then provided that the system of equations (1.6) possesses non-singular real and p-adic solutions for each prime number p, one has*

$$N(B) \sim \sigma_\infty \left(\prod_p \sigma_p \right) B^{sd-K}. \tag{1.7}$$

In particular, the system (1.6) satisfies the Hasse Principle.

We note that [21, Theorem 9.1] delivers the same conclusion as Theorem 1.4 in the special case $d = 1$. As is apparent from the lower bound supplied by Theorem 1.2 and the ensuing discussion, there exist choices of coefficients $\mathbf{a}$ for which the asymptotic formula (1.7) necessarily fails when k is large and

$$s \leqslant \frac{d}{d+1}(1 + O(d/k))r(k+1). \tag{1.8}$$

Consequently, the bound on the number of variables in the hypotheses of Theorem 1.4 is within a factor $2 + 2/d + O(d/k)$ of the best possible bound for such systems. Indeed, the argument underlying the proof of Theorem 1.2 shows that such remains true in wider generality. The point here is that special subvarieties contain the bulk of the set of integral solutions whenever the bound (1.8) holds on the number s of blocks of d variables. Conclusions available hitherto of the type presented in Theorem 1.4 impose bounds on the number of blocks of variables weaker than our own by a factor of order $\log(rk)$.

As a special case of Corollary 11.2, we obtain an asymptotic formula for the mean value $J_{s,k,d}(X)$.

Theorem 1.5. *Let k and d be natural numbers with $k \geqslant 2$. Then whenever $s \geqslant r(k+1)+1$, there exist positive constants $\mathscr{C} = \mathscr{C}(s,k,d)$ and $\delta = \delta(k,d)$ such that*

$$J_{s,k,d}(X) = \mathscr{C} X^{2sd-K} + O(X^{2sd-K-\delta}).$$

A conclusion analogous to that of Theorem 1.5 is obtained in [9, Theorem 1.3], subject to the condition that k be sufficiently large and

$$s \geqslant rk(\tfrac{2}{3}\log r + \tfrac{1}{2}\log k + \log\log k + 2d + 4).$$

Again, the conclusion of Theorem 1.5 is much superior. When $d = 1$ and $k \geqslant 3$, meanwhile, the conclusion of Theorem 1.5 is a consequence of [21, Theorem 1.2].

As a penultimate application of the bounds supplied by Theorem 1.1, in §11 we consider the rational linear spaces of projective dimension $d-1$ lying on the diagonal hypersurface

$$c_1z_1^k+\ldots+c_sz_s^k=0, \tag{1.9}$$

with $c_1,\ldots,c_s$ fixed non-zero integers. Such a linear space may be written in the form

$$\mathscr{L}(\mathbf{x}_1,\ldots,\mathbf{x}_d)=\{t_1\mathbf{x}_1+\ldots+t_d\mathbf{x}_d : t_1,\ldots,t_d\in\mathbb{Q}\},$$

for suitable linearly independent vectors $\mathbf{x}_1,\ldots,\mathbf{x}_d\in\mathbb{Z}^s$. As noted in [9], by substituting into (1.9) and using the multinomial theorem to collect together coefficients of $\mathbf{t}^{\mathbf{i}}$, one finds that the linear space $\mathscr{L}(\overline{\mathbf{x}})$ corresponds to a solution $\mathbf{y}_1,\ldots,\mathbf{y}_s\in\mathbb{Z}^d$ of the Diophantine system

$$c_1\mathbf{y}_1^{\mathbf{i}}+\ldots+c_s\mathbf{y}_s^{\mathbf{i}}=0\quad(|\mathbf{i}|=k). \tag{1.10}$$

This correspondence is made explicit by means of the simple relation

$$x_{ij}=y_{ji}\quad(1\leqslant i\leqslant d,\,1\leqslant j\leqslant s).$$

Write $N_{s,k,d}(X)$ for the number of integral solutions of the system (1.10) with $|\overline{\mathbf{y}}|\leqslant X$, and put

$$L=k\binom{k+d-1}{k}.$$

In §11 we indicate how to prove an asymptotic formula for $N_{s,k,d}(X)$ subject to the condition that $s\geqslant 2r(k+1)+1$. In this context, we say that the integral s-tuple $\mathbf{c}$ is a *non-singular choice of coefficients for k and d* when the system of equations (1.10) has non-singular real and p-adic solutions, for every prime number p.

Theorem 1.6. *Suppose that s, k and d are natural numbers with $k\geqslant 2$ and $s\geqslant 2r(k+1)+1$. Suppose further that $\mathbf{c}\in(\mathbb{Z}\setminus\{\mathbf{0}\})^s$ is a non-singular choice of coefficients for k and d. Then there exist positive constants $\mathscr{D}=\mathscr{D}(s,k,d;\mathbf{c})$ and $\nu=\nu(k,d)$ such that*

$$N_{s,k,d}(X)=\mathscr{D}X^{sd-L}+O(X^{sd-L-\nu}).$$

In particular, one finds that whenever $s\geqslant 2r(k+1)+1$ and appropriate local solubility conditions are met, then the hypersurface defined by (1.9) contains an abundance of rational linear spaces of projective dimension $d-1$. A perusal of [9] reveals that, for sufficiently large values of k, a similar conclusion is asserted by Theorem 1.4 of the latter source, subject instead to the more stringent condition

$$s\geqslant rk(\tfrac{4}{3}\log r+\log k+2\log\log k+4d+8).$$

We remark that the lower bound $s\geqslant 2r(k+1)+1$ in Theorem 1.6 should be susceptible to some small improvement by adapting the methods of [20] to the present multidimensional setting. Moreover, when the degree k is very small, an approach of the first author [10] motivated by a method of Hua proves superior in some situations.

Further applications within the orbit of our methods and bounds include the generalised Waring problem of representing a given polynomial

$$\Phi(t_1, \dots, t_d) \in \mathbb{Z}[\mathbf{t}]$$

in the form

$$\Phi(\mathbf{t}) = \sum_{j=1}^{s} (x_{1j}t_1 + \dots + x_{dj}t_d)^k,$$

and also results concerning the number of integral solutions of Diophantine inequalities modulo 1. We refer the reader to [2] for a discussion of some such problems, and leave to the reader the satisfaction of incorporating our new bounds into the established methods so as to make similarly striking improvements over the previous state of knowledge.

As a final application of our new bounds for multidimensional Weyl sums, we announce an application in additive combinatorics based on the second author's recent work [11] on translation invariant systems of equations devoid of solutions in multidimensional sets. In Theorem 11.3 below we present a conclusion more general than the one we presently record in Theorem 1.7. For the purpose at hand, we describe the integral s-tuple $\mathbf{c}$ as an *extended non-singular choice of coefficients for k and d* when (i) one has $c_1 + \dots + c_s = 0$, and (ii) the system of equations

$$c_1\mathbf{y}_1^{\mathbf{i}} + \dots + c_s\mathbf{y}_s^{\mathbf{i}} = 0 \quad (1 \leqslant |\mathbf{i}| \leqslant k) \tag{1.11}$$

has non-singular real and p-adic solutions, for every prime number p.

Certain solutions of the system (1.11) are atypically simple to obtain, such as the trivial solutions lying on the diagonal $\mathbf{y}_1 = \mathbf{y}_2 = \dots = \mathbf{y}_s$. We formalise this notion by distinguishing two types of special solutions $\overline{\mathbf{y}}$ of (1.11). We describe $\overline{\mathbf{y}}$ as *projected* when there is a translate of a proper subspace of $\mathbb{Q}^d$ that contains all of $\mathbf{y}_1, \dots, \mathbf{y}_s$. The aforementioned diagonal solutions are therefore projected, since they lie in a translate of the trivial subspace $\{\mathbf{0}\}$ of $\mathbb{Q}^d$. Also, we say that $\overline{\mathbf{y}}$ is a *subset-sum solution* when there exists a partition $\{1, 2, \dots, s\} = \mathscr{J}_1 \cup \dots \cup \mathscr{J}_l$, into $l \geqslant 2$ disjoint non-empty sets $\mathscr{J}_v$, such that for $1 \leqslant v \leqslant l$ one has

$$\sum_{u \in \mathscr{J}_v} c_u\mathbf{y}_u^{\mathbf{i}} = 0 \quad (1 \leqslant |\mathbf{i}| \leqslant k).$$

In the special case in which

$$\sum_{u \in \mathscr{J}_v} c_u = 0 \quad (1 \leqslant v \leqslant l),$$

one sees that there are trivial subset-sum solutions in which $\mathbf{y}_u = \mathbf{y}_w$ whenever $u \in \mathscr{J}_v$ and $w \in \mathscr{J}_v$ $(1 \leqslant v \leqslant l)$.

Theorem 1.7. *Suppose that s, k and d are natural numbers with $k \geqslant 2$ and $s \geqslant 2r(k+1)+1$. Suppose further that $\mathbf{c} \in (\mathbb{Z} \setminus \{\mathbf{0}\})^s$ is an extended non-singular choice of coefficients for k and d, so that $c_1 + \dots + c_s = 0$. Let $\mathscr{A}$*

be a subset of $\mathbb{Z}^d \cap [1,N]^d$*, and suppose that the only solutions of the system (1.11) from* $\mathscr{A}$ *are either projected or subset-sum solutions. Then one has*

$$\mathrm{card}(\mathscr{A}) \ll N^d(\log\log N)^{-1/(s-1)}.$$

This theorem is a higher dimensional cousin of [11, Theorem 5.1], which supplies an analogous conclusion for a case involving binary forms. Theorem 1.7 shows that when $\mathrm{card}(\mathscr{A})$ grows more rapidly than $N^d(\log\log N)^{-1/(s-1)}$, then the system (1.11) contains solutions from $\mathscr{A}$ besides such obvious ones as the diagonal solutions with $\mathbf{y}_1 = \ldots = \mathbf{y}_s$. As we see in §3, the extended system (1.11) contains more general special subvarieties defined by means of a projection process, the simplest of which set one or more variables to be zero. In Theorem 11.3 we present a conclusion that refines Theorem 1.7 in which, under the same hypotheses concerning the cardinality of $\mathscr{A}$, one finds that the system (1.11) contains solutions from $\mathscr{A}$ which avoid all of these special subvarieties. In this way, one may legitimately describe the solutions of (1.11) thus shown to exist as honestly non-trivial. The interested reader will find the necessary ideas in earlier work [11] of the second author.

It may be useful to provide an informal sketch hinting at the argument underlying the proof of Theorem 1.1 so that the reader is better prepared to draw parallels with previous approaches. A more comprehensively illuminated sketch of this argument in the case $d = 1$ may be found in [21, §2]. In common with the previous approaches of [2] and [9], the basic tool employed in our proof of Theorem 1.1 is a (so-called) p-adic iteration mirroring the one devised by Linnik [6] in the classical setting with $d = 1$. Thus, we begin by artificially introducing a congruence condition, modulo a suitable prime p, amongst the bulk of the variables underlying the mean value (1.5). An application of Hölder's inequality leads to a new mean value in which the latter variables lie in common congruence classes across blocks. At this point, the multiple translation invariance of the system (1.1) may be utilised so as to pass to the zero congruence class, and thereby a congruence condition is forced on a subset of the variables of greater strength than that previously introduced. The approach of [2] is to choose the prime p in such a way that this strong congruence condition forces a diagonal condition amongst blocks of variables, and thereby one is able to bound a mean value involving $2(s+r)$ blocks of d variables in terms of a corresponding mean value involving $2s$ blocks of d variables. In [9] the strong congruence condition is interpreted as a differencing process analogous to, though more efficient than, that of Weyl. By appropriate use of the Cauchy-Schwarz inequalities, one is able to repeat this *efficient differencing* process, deferring the moment at which to force the diagonal condition. In the present paper, following [21], we instead interpret the strong congruence condition as an efficient method of imposing a second artificial congruence condition amongst variables. By appropriate application of Hölder's inequality, one recovers a new mean value resembling that obtained in the first step, but now yielding a fresh congruence condition amongst variables significantly stronger than before. If one begins with a mean value significantly larger in size than anticipated, then repeated application of this *efficient congruencing*

procedure yields a related mean value larger in size than that anticipated by an amount so large that even a trivial estimate demonstrates the presumed initial deviation from the expected size to be untenable. In this way, one shows that the mean value under consideration has size very close to that expected.

We finish by emphasising that the methods of this paper are robust to changes of the ambient ring. Thus the rational integers $\mathbb{Z}$ central to this paper may be replaced with the ring of integers from a number field, or the polynomial ring $\mathbb{F}_q[t]$, without diminishing the strength of the ensuing estimates. Such ideas have been explored very recently in the case $d = 1$ in work emerging from the body of research exploiting the efficient congruencing method (see [7] and [23]).

In §2 we introduce the general translation-dilation invariant systems which constitute the central objects of attention in this paper. Then, in §3, we discuss the lower bounds recorded in Theorem 1.2. The notation and infrastructure required for our most general conclusions is discussed in §4, and then in §5 we derive the basic mean value estimates which initiate our efficient congruencing argument. Next, in §6, we provide estimates for the number of solutions of a system of basic congruences. Here, the singular locus of the system is of particular concern. The conditioning process, required to guarantee appropriate non-singularity conditions, is engineered in §7, and in §8 we discuss the efficient congruencing process itself. In §9 we combine the output of §§7 and 8 so as to deliver Theorem 1.1 via an iterative process. Consequences for Weyl-type estimates are discussed in §10, yielding the conclusion of Theorem 1.3. Finally, in §11 we sketch the arguments required to establish the Diophantine consequences recorded in Theorems 1.4-1.7.

2. Translation-dilation invariant systems

In order to describe our most general conclusions, we must introduce some notation having flexibility sufficient for our needs. An overly prescriptive approach has the potential to shroud the details of our arguments in a thick blanket of impenetrable symbols. With this undesirable potential outcome in mind, we opt for a somewhat abstract approach, and only later do we spend time detailing the most interesting situations.

Let r, s and d be natural numbers, and consider a system of homogeneous polynomials $\mathbf{F} = (F_1, \dots, F_r)$, where $F_j(\mathbf{z}) \in \mathbb{Z}[z_1, \dots, z_d]$ $(1 \leqslant j \leqslant r)$. We investigate the system of Diophantine equations

$$\sum_{i=1}^{s} (\mathbf{F}(\mathbf{x}_i) - \mathbf{F}(\mathbf{y}_i)) = \mathbf{0}, \tag{2.1}$$

in which $\mathbf{x}_i = (x_{i1}, \dots, x_{id})$ and $\mathbf{y}_i = (y_{i1}, \dots, y_{id})$ for $1 \leqslant i \leqslant s$. Note that, in view of our conventions concerning vector notation, the system (2.1) consists of r simultaneous Diophantine equations. Write $\overline{\mathbf{x}} = (\mathbf{x}_1, \dots, \mathbf{x}_s)$ and $\overline{\mathbf{y}} = (\mathbf{y}_1, \dots, \mathbf{y}_s)$, and denote by $J_s(X; \mathbf{F})$ the number of integral solutions of the system (2.1) with $1 \leqslant \overline{\mathbf{x}}, \overline{\mathbf{y}} \leqslant X$.

In this paper we are concerned with translation-dilation invariant systems of the shape (2.1). With the discussion to come in mind, we take a pragmatic approach to defining such systems. We say that the system $\mathbf{F} = (F_1, \dots, F_r)$ is *translation-dilation invariant* if:
(i) the polynomials $F_1, \dots, F_r$ are each homogeneous of positive degree, and
(ii) there exist polynomials

$$c_{jl} \in \mathbb{Z}[\xi_1, \dots, \xi_d] \quad (1 \leqslant j \leqslant r \text{ and } 0 \leqslant l \leqslant j),$$

with $c_{jj} = 1$ for $1 \leqslant j \leqslant r$, having the property that whenever $\boldsymbol{\xi} \in \mathbb{Z}^d$, then

$$F_j(\mathbf{x} + \boldsymbol{\xi}) = c_{j0}(\boldsymbol{\xi}) + \sum_{l=1}^{j} c_{jl}(\boldsymbol{\xi}) F_l(\mathbf{x}) \quad (1 \leqslant j \leqslant r). \tag{2.2}$$

Extend the definition of the coefficients c_{jl} by putting $c_{jl}(\boldsymbol{\xi}) = 0$ when $l > j$. Then on writing $\mathbf{c}_0(\boldsymbol{\xi}) = (c_{j0}(\boldsymbol{\xi}))_{1 \leqslant j \leqslant r}$ and $C(\boldsymbol{\xi})$ for the matrix $(c_{jl}(\boldsymbol{\xi}))_{1 \leqslant j,l \leqslant r}$, we see that the relations (2.2) are summarised by the formula

$$\mathbf{F}(\mathbf{x} + \boldsymbol{\xi}) = C(\boldsymbol{\xi})\mathbf{F}(\mathbf{x}) + \mathbf{c}_0(\boldsymbol{\xi}). \tag{2.3}$$

Notice that the matrix $C(\boldsymbol{\xi})$ is lower unitriangular, which is to say that it is a lower triangular matrix whose main diagonal entries are all 1. Suppose that s is a natural number, that λ is a non-zero rational number, and $\boldsymbol{\xi} \in \mathbb{Z}^d$. Then we see from (2.2) that the Diophantine system (2.1) possesses an integral solution $\mathbf{x}, \mathbf{y}$ if and only if one has

$$\sum_{i=1}^{s} (\mathbf{F}(\lambda \mathbf{x}_i + \boldsymbol{\xi}) - \mathbf{F}(\lambda \mathbf{y}_i + \boldsymbol{\xi})) = \mathbf{0}. \tag{2.4}$$

This observation justifies the description of such systems of equations as translation-dilation invariant. We should note that while this formal definition facilitates many of our arguments, it is clear that one may rearrange the ordering of the forms, and also consider independent linear combinations of the original forms, without altering the number of integral solutions of the system (2.1) counted by $J_s(X; \mathbf{F})$. Thus we may be expedient in most circumstances, and instead describe a system as translation-dilation invariant when it is equivalent in such a manner to some new system which is translation-dilation invariant in the strict sense.

We emphasise that translation-dilation invariant systems are easily generated. Given a collection of homogeneous polynomials

$$G_1, \dots, G_h \in \mathbb{Z}[z_1, \dots, z_d],$$

consider the set $\mathscr{F}$ consisting of all the partial derivatives

$$\frac{\partial^{l_1 + \dots + l_d} G_j(\mathbf{z})}{\partial z_1^{l_1} \dots \partial z_d^{l_d}} \quad (1 \leqslant j \leqslant h), \tag{2.5}$$

with $l_i \geqslant 0$ $(1 \leqslant i \leqslant d)$. Plainly, when $l_1 + \dots + l_d$ exceeds the largest total degree of any of the polynomials G_j, this partial derivative vanishes. The set $\mathscr{F}$ is consequently finite. Let $\mathscr{F}_0$ denote the subset of $\mathscr{F}$ consisting of all polynomials in $\mathscr{F}$ having positive degree. We write $\mathscr{F}_0 = \{F_1, \dots, F_r\}$,

labelling the elements in such a way that $\deg F_1 \leqslant \deg F_2 \leqslant \ldots \leqslant \deg F_r$. An application of the multidimensional version of Taylor's theorem now shows that the relations (2.2) hold for some choice of coefficients $c_{jl}(\boldsymbol{\xi}) \in \mathbb{Z}[\xi_1, \ldots, \xi_d]$ satisfying $c_{jj}(\boldsymbol{\xi}) = 1$ $(1 \leqslant j \leqslant r)$. Since we may replace the set of forms $\mathscr{F}_0$ by any subset whose span contains the polynomials $F_1, \ldots, F_r$, there is no loss of generality in supposing the set $\{F_1, \ldots, F_r\}$ to be linearly independent. Such a system of forms we call *reduced*.

Finally, by replacing the forms $F_1, \ldots, F_r$ by appropriate linear combinations of the original forms, we find that there is no loss of generality also in supposing that the matrix $C(\boldsymbol{\xi})$ with entries $c_{jl}(\boldsymbol{\xi})$ is lower unitriangular. This new system $\mathbf{F} = (F_1, \ldots, F_r)$, generated from the partial derivatives (2.5), is a reduced translation-dilation invariant system.

We are now almost equipped to state our main theorem, but first pause to introduce some parameters associated with a translation-dilation invariant system of polynomials $\mathbf{F}$. When $\mathbf{F} = (F_1, \ldots, F_r)$ consists of polynomials $F_j(\mathbf{z}) \in \mathbb{Z}[z_1, \ldots, z_d]$, we refer to the number of variables $d = d(\mathbf{F})$ in $\mathbf{F}$ as the *dimension* of the system. In addition, we describe the number of forms $r = r(\mathbf{F})$ comprising $\mathbf{F}$ as the *rank* of the system. We write $k_j = k_j(\mathbf{F})$ for the total degree of the polynomial F_j, and then define the *degree* $k = k(\mathbf{F})$ of the system $\mathbf{F}$ by

$$k(\mathbf{F}) = \max_{1 \leqslant j \leqslant r} k_j(\mathbf{F}),$$

and the *weight* $K = K(\mathbf{F})$ by

$$K(\mathbf{F}) = \sum_{j=1}^{r} k_j(\mathbf{F}).$$

Our goal in §§4-9 is the proof of the following mean value estimate, which represents the main theorem of this paper.

Theorem 2.1. *Let $\mathbf{F}$ be a reduced translation-dilation invariant system of polynomials having dimension d, rank r, degree k and weight K. Suppose that s is a natural number with $s \geqslant r(k+1)$. Then for each $\varepsilon > 0$, one has $J_s(X; \mathbf{F}) \ll X^{2sd-K+\varepsilon}$.*

So far as we are aware, no mean value estimate available in the literature has generality to compete with Theorem 2.1. Moreover, when $d \geqslant 2$ the estimates available hitherto are considerably weaker, even in the special situations in which they are applicable. In order to illustrate the ease with which estimates may be extracted from Theorem 2.1, we finish this section with a brief discussion of some simple cases, and in particular we show how to establish Theorem 1.1 as a consequence of Theorem 2.1.

(a) The classical system of Vinogradov [16], [17]. Consider the seed polynomial z^k $(k \geqslant 1)$. By taking successive derivatives, we find that an associated reduced translation-dilation invariant system of polynomials is $\mathbf{F} = (z^k, z^{k-1}, \ldots, z)$.

This system has dimension 1, rank k, degree k and weight

$$K = \sum_{j=1}^{k} j = \tfrac{1}{2}k(k+1).$$

Then it follows from Theorem 2.1 that when $s \geqslant k(k+1)$, one has

$$J_s(X; \mathbf{F}) \ll X^{2s-\frac{1}{2}k(k+1)+\varepsilon}.$$

This estimate recovers the main conclusion of the third author's recent work introducing the efficient congruencing method to Vinogradov's mean value theorem (see [21, Theorem 1.1]). We note that subsequent work of the third author leads to the improved constraint $s \geqslant k^2 - 1$ on the number of variables in this conclusion (see [22, Theorem 1.1]).

(b) The system of Parsell [9]. Consider the situation with $d \geqslant 2$ and seed polynomials $z_1^{l_1} z_2^{l_2} \dots z_d^{l_d}$ ($|\mathbf{l}| = k$). By taking successive partial derivatives, we find that an associated reduced translation-dilation invariant system of polynomials is

$$\mathbf{F} = (z_1^{i_1} z_2^{i_2} \dots z_d^{i_d} : 1 \leqslant |\mathbf{i}| \leqslant k).$$

This system has dimension d, rank

$$r = \sum_{1 \leqslant |\mathbf{i}| \leqslant k} 1 = \binom{k+d}{d} - 1, \tag{2.6}$$

degree k and weight

$$K = \sum_{l=1}^{k} l \sum_{|\mathbf{i}|=l} 1 = \sum_{l=1}^{k} l \binom{l+d-1}{l} = \frac{d}{d+1}(r+1)k. \tag{2.7}$$

In this instance, it follows from Theorem 2.1 that when $s \geqslant r(k+1)$, one has $J_s(X; \mathbf{F}) \ll X^{2sd-K+\varepsilon}$. In view of (1.2) and (1.3), this completes the proof of Theorem 1.1.

(c) The system of Arkhipov, Karatsuba and Chubarikov [2]. Consider the situation with $d \geqslant 2$ and $l \geqslant 1$ and the seed polynomial $z_1^l z_2^l \dots z_d^l$. By taking successive partial derivatives, we find that an associated reduced translation-dilation invariant system of polynomials is

$$\mathbf{F} = (z_1^{i_1} z_2^{i_2} \dots z_d^{i_d} : 0 \leqslant \mathbf{i} \leqslant l,\ \mathbf{i} \neq \mathbf{0}). \tag{2.8}$$

This system has dimension d, rank

$$r = \sum_{\substack{0 \leqslant \mathbf{i} \leqslant l \\ \mathbf{i} \neq \mathbf{0}}} 1 = \sum_{0 \leqslant i_1 \leqslant l} \dots \sum_{0 \leqslant i_d \leqslant l} 1 - 1 = (l+1)^d - 1, \tag{2.9}$$

degree dl, and weight

$$\begin{aligned} K &= \sum_{0\leqslant \mathbf{i}\leqslant l} |\mathbf{i}| = \sum_{0\leqslant i_1\leqslant l} \dots \sum_{0\leqslant i_d\leqslant l} (i_1+\dots+i_d) \\ &= d \sum_{0\leqslant i_1\leqslant l} \dots \sum_{0\leqslant i_d\leqslant l} i_d \\ &= d(l+1)^{d-1}\cdot \tfrac{1}{2}l(l+1) = \tfrac{1}{2}dl(l+1)^d. \end{aligned} \tag{2.10}$$

In this instance, Theorem 2.1 delivers a conclusion important enough to summarise as a corollary.

Corollary 2.2. *Let d and l be natural numbers, and let $\mathbf{F}$ be the reduced translation-dilation invariant system given by (2.8). Suppose that s is a natural number with $s \geqslant (dl+1)((l+1)^d-1)$. Then for each $\varepsilon > 0$, one has*

$$J_s(X;\mathbf{F}) \ll X^{2sd-K+\varepsilon},$$

where $K = \frac{1}{2}dl(l+1)^d$.

A conclusion similar to that provided by Corollary 2.2, but with the condition $s \geqslant (dl+1)((l+1)^d-1)$ replaced by

$$s \geqslant Cdl(l+1)^d \log(dl(l+1)^d),$$

for a suitable positive constant C, may be extracted from [2, Theorem 1 of Chapter III.1]. The superiority of our new bound is self-evident.

(d) Simple binary systems. A system of relevance to recent work in quantitative arithmetic geometry (see [14, §4.15],[15]) deserves to be singled out for special attention. Consider the situation with $k_1 \geqslant k_2 \geqslant 1$ and the seed polynomial $z_1^{k_1}z_2^{k_2}$. By taking successive partial derivatives, we find that an associated reduced translation-dilation invariant system of polynomials is

$$\mathbf{F} = (z_1^{i_1}z_2^{i_2} : 0 \leqslant i_1 \leqslant k_1,\ 0 \leqslant i_2 \leqslant k_2 \text{ and } (i_1,i_2) \neq (0,0)). \tag{2.11}$$

This system has dimension 2, rank

$$r = \sum_{0\leqslant i_1\leqslant k_1} \sum_{0\leqslant i_2\leqslant k_2} 1 - 1 = (k_1+1)(k_2+1)-1,$$

degree k_1+k_2, and weight

$$\begin{aligned} K &= \sum_{0\leqslant i_1\leqslant k_1} \sum_{0\leqslant i_2\leqslant k_2} (i_1+i_2) \\ &= (k_1+1)\cdot \tfrac{1}{2}k_2(k_2+1) + (k_2+1)\cdot \tfrac{1}{2}k_1(k_1+1) \\ &= \tfrac{1}{2}(k_1+k_2)(k_1+1)(k_2+1). \end{aligned}$$

By applying Theorem 2.1, we obtain the following corollary.

Corollary 2.3. *Let $k_1, k_2 \in \mathbb{N}$, and let $\mathbf{F}$ be the reduced translation-dilation invariant system given by (2.11). Suppose that s is a natural number with $s \geqslant (k_1k_2+k_1+k_2)(k_1+k_2+1)$. Then for each $\varepsilon > 0$, one has*

$$J_s(X;\mathbf{F}) \ll X^{4s-K+\varepsilon},$$

where $K = \frac{1}{2}(k_1+k_2)(k_1+1)(k_2+1)$.

(e) The binary systems of Prendiville [11]. Consider the situation with $k \geqslant 1$ and the seed polynomial given by the binary form $\Phi(z_1, z_2) \in \mathbb{Z}[z_1, z_2]$ of degree k. In this instance, we extract the partial derivatives

$$\frac{\partial^{i_1+i_2}\Phi(z_1, z_2)}{\partial z_1^{i_1}\partial z_2^{i_2}} \quad (i_1 \geqslant 0,\, i_2 \geqslant 0),$$

and restrict attention to any subset which spans the set of all partial derivatives of positive degree, yet is linearly independent over $\mathbb{Q}$. We take the polynomials in this spanning set to be our reduced translation-dilation invariant system $\mathbf{F}$. The number of partial derivatives with $i_1 + i_2 = l$ is plainly $l + 1$, while the number of monomials $z_1^{j_1} z_2^{j_2}$ with $j_1 + j_2 = m$ is $m + 1$. Thus we see that this system has dimension $d = 2$, rank

$$r \leqslant \sum_{l=0}^{[k/2]}(l+1) + \sum_{m=1}^{k-[k/2]-1}(m+1) \leqslant \tfrac{1}{4}k(k+4),$$

degree k and weight

$$K \leqslant \sum_{l=0}^{[k/2]}(k-l)(l+1) + \sum_{m=1}^{k-[k/2]-1} m(m+1) \leqslant \tfrac{1}{8}k(k+2)^2.$$

In typical situations, indeed, one has $K \sim \frac{1}{8}k^3$. By applying Theorem 2.1, we deduce that when $s \geqslant \frac{1}{4}k(k+1)(k+4)$, one has $J_s(X; \mathbf{F}) \ll X^{4s-K+\varepsilon}$. This conclusion may be compared with the mean value estimate underlying [11, Theorem 1.3], which delivers a similar conclusion for $s \geqslant (\frac{3}{8} + o(1))k^3 \log k$. The constraint on s imposed in our present work is therefore stronger by a factor $(\frac{3}{2} + o(1)) \log k$.

3. Lower bounds

In order to put into perspective the upper bounds recorded in Theorem 2.1, and such corollaries as Theorem 1.1, we consider in this section the topic of lower bounds for the mean value $J_s(X; \mathbf{F})$. Here one must consider integral solutions to the system of equations (2.1) of two types. On the one hand, there are typical solutions whose contribution to $J_s(X; \mathbf{F})$ we expect to be given by a product of local densities. On the other hand, there are integral solutions lying on special subvarieties, the most obvious of which are diagonal linear spaces such as that given by $\mathbf{x}_i = \mathbf{y}_i$ $(1 \leqslant i \leqslant s)$. It transpires that when $d > 1$, there are special subvarieties not of the latter type which potentially make the dominant contribution to $J_s(X; \mathbf{F})$. In order to describe the latter subvarieties, we must introduce some further notation.

Let r and d be natural numbers, and consider a system of translation-dilation invariant polynomials $\mathbf{F} = (F_1, \dots, F_r)$, where $F_j(\mathbf{z}) \in \mathbb{Z}[z_1, \dots, z_d]$ $(1 \leqslant j \leqslant r)$. Let δ be a natural number with $1 \leqslant \delta \leqslant d - 1$, and consider indices i_l $(1 \leqslant l \leqslant \delta)$ satisfying

$$1 \leqslant i_1 < i_2 < \ldots < i_\delta \leqslant d. \tag{3.1}$$

We say that the system of polynomials $\mathbf{G} = (G_1, \dots, G_r)$, where $G_j(\mathbf{w}) \in \mathbb{Z}[w_1, \dots, w_\delta]$ $(1 \leqslant j \leqslant r)$, is the *orthogonal projection of* $\mathbf{F}$ *determined by* $\mathbf{i}$ when

$$G_j(\mathbf{w}) = F_j(\boldsymbol{\zeta}) \quad (1 \leqslant j \leqslant r),$$

in which $\zeta_m = w_l$ when $m = i_l$ for some index l with $1 \leqslant l \leqslant \delta$, and $\zeta_m = 0$ when $m \notin \{i_1, \dots, i_\delta\}$. The system $\mathbf{G}$ remains translation-dilation invariant, and may be replaced by an equivalent reduced system $\mathbf{G}'$. We describe $\mathbf{G}'$ as a *reduced orthogonal projection of* $\mathbf{F}$ *determined by* $\mathbf{i}$. Finally, write $\pi_\delta(\mathbf{F})$ for the set of all reduced orthogonal projections of $\mathbf{F}$ determined by sets of indices $\{i_1, \dots, i_\delta\}$ satisfying (3.1). We remark that these orthogonal projections are in fact a special case of the more general projections introduced in the preamble to Theorem 1.7. In this section we consider only the former projections, since they are simpler to analyse and in any case deliver all of the salient features of importance for our discussion of lower bounds.

In order to facilitate our subsequent discussion, we define the polynomial $\psi(x; \boldsymbol{\alpha}) = \psi(x; \boldsymbol{\alpha}; \mathbf{F})$ by putting

$$\psi(\mathbf{x}; \boldsymbol{\alpha}; \mathbf{F}) = \sum_{i=1}^{r} \alpha_i F_i(\mathbf{x}), \tag{3.2}$$

and then define the associated exponential sum $f(\boldsymbol{\alpha}) = f(\boldsymbol{\alpha}; X; \mathbf{F})$ by

$$f(\boldsymbol{\alpha}; X; \mathbf{F}) = \sum_{1 \leqslant \mathbf{x} \leqslant X} e(\psi(\mathbf{x}; \boldsymbol{\alpha}; \mathbf{F})). \tag{3.3}$$

By orthogonality, we then have

$$J_s(X; \mathbf{F}) = \oint |f(\boldsymbol{\alpha}; X; \mathbf{F})|^{2s} \, \mathrm{d}\boldsymbol{\alpha}. \tag{3.4}$$

We are now equipped to describe our most general lower bound for the mean value $J_s(X; \mathbf{F})$.

Theorem 3.1. *Let* $\mathbf{F}$ *be a reduced translation-dilation invariant system of polynomials having dimension* d *and weight* K*. Then for each natural number* s*, one has*

$$J_s(X; \mathbf{F}) \gg X^{sd} + X^{2sd-K} + \sum_{\delta=1}^{d-1} X^{d-\delta} \max_{\mathbf{G} \in \pi_\delta(\mathbf{F})} J_s(X; \mathbf{G}).$$

Proof. We consider first typical solutions of the system (2.1) not constrained to lie on special subvarieties. Suppose that $\mathbf{F}$ has rank r, and write k_j for the degree of F_j for $1 \leqslant j \leqslant r$. There exists a positive number A, depending at most on d, k and the coefficients of $\mathbf{F}$, such that whenever $1 \leqslant \mathbf{x} \leqslant X$, one has

$$|F_j(\mathbf{x})| \leqslant AX^{k_j} \quad (1 \leqslant j \leqslant r).$$

Consequently, when $1 \leqslant \overline{\mathbf{x}}, \overline{\mathbf{y}} \leqslant X$, one sees that for $1 \leqslant j \leqslant r$ the integer

$$\sum_{l=1}^{s} (F_j(\mathbf{x}_l) - F_j(\mathbf{y}_l))$$

lies in the interval $[-2sAX^{k_j}, 2sAX^{k_j}]$. We therefore deduce by means of orthogonality in combination with the triangle inequality and (3.4) that

$$[X]^{2sd} = \sum_{\substack{|h_j|\leqslant 2sAX^{k_j}\\ (1\leqslant j\leqslant r)}} \oint |f(\boldsymbol{\alpha}; X; \mathbf{F})|^{2s} e(-\alpha_1 h_1 - \ldots - \alpha_r h_r)\, \mathrm{d}\boldsymbol{\alpha}$$
$$\ll \Big(\prod_{1\leqslant j\leqslant r} X^{k_j} \Big) \oint |f(\boldsymbol{\alpha}; X; \mathbf{F})|^{2s}\, \mathrm{d}\boldsymbol{\alpha} = X^K J_s(X; \mathbf{F}).$$

Thus we conclude that

$$J_s(X; \mathbf{F}) \gg X^{2sd-K}. \tag{3.5}$$

Next, by considering the diagonal solutions of (2.1) with $1 \leqslant \overline{\mathbf{x}}, \overline{\mathbf{y}} \leqslant X$ and $\mathbf{x}_j = \mathbf{y}_j$ $(1 \leqslant j \leqslant s)$, we obtain the lower bound

$$J_s(X; \mathbf{F}) \gg X^{sd}. \tag{3.6}$$

We now come to consider the solutions of the system (2.1) lying on certain additional special subvarieties. We assert that when $\mathbf{G}$ is a reduced translation-dilation invariant system with dimension $e \geqslant 2$, then for every system $\mathbf{H}$ lying in $\pi_{e-1}(\mathbf{G})$, one has

$$J_s(X; \mathbf{G}) \gg X J_s(X; \mathbf{H}). \tag{3.7}$$

In view of (3.5) and (3.6), the lower bound claimed in the statement of Theorem 3.1 then follows by induction.

Let $\mathbf{G} = (G_1, \ldots, G_u)$ be a reduced translation-dilation invariant system of dimension $e > 1$, and consider a system $\mathbf{H} \in \pi_{e-1}(\mathbf{G})$. The system $\mathbf{H}$ is the orthogonal projection of $\mathbf{G}$ determined by some $(e-1)$-tuple $\mathbf{i} = (i_1, \ldots, i_{e-1})$. By relabelling variables, if necessary, we may suppose that $\mathbf{i} = (1, 2, \ldots, e-1)$. Consider now the system of equations

$$\sum_{i=1}^{s} (\mathbf{G}(\mathbf{x}_i) - \mathbf{G}(\mathbf{y}_i)) = \mathbf{0}. \tag{3.8}$$

Let a be an integer with $1 \leqslant a \leqslant X$, set $\mathbf{a} = (0, \ldots, 0, a)$, and consider the effect of the translation $(\mathbf{x}_i, \mathbf{y}_i) \longmapsto (\mathbf{x}_i - \mathbf{a}, \mathbf{y}_i - \mathbf{a})$. In view of the translation invariance of the system (3.8) that is a consequence of the discussion leading to (2.4), one finds that whenever the system of equations

$$\sum_{i=1}^{s} (\mathbf{G}(\mathbf{x}_i - \mathbf{a}) - \mathbf{G}(\mathbf{y}_i - \mathbf{a})) = \mathbf{0} \tag{3.9}$$

is satisfied, then so too is the system (3.8). If we substitute $x_{ie} = y_{ie} = a$ $(1 \leqslant i \leqslant s)$, then we find that $\mathbf{G}(\mathbf{x}_i - \mathbf{a}) = \mathbf{H}(\mathbf{x}_i)$, where $\mathbf{H}$ is the aforementioned orthogonal projection of $\mathbf{G}$ determined by $(1, 2, \ldots, e-1)$. Write $\mathbf{H}'$ for any reduced system equivalent to $\mathbf{H}$. Then we conclude that whenever $\overline{\mathbf{z}}, \overline{\mathbf{w}}$ is a solution of the system

$$\sum_{i=1}^{s} (\mathbf{H}'(\mathbf{z}_i) - \mathbf{H}'(\mathbf{w}_i)) = \mathbf{0},$$

then the system (3.9) has the solution

$$\mathbf{x}_i = (\mathbf{z}_i, a) \quad \text{and} \quad \mathbf{y}_i = (\mathbf{w}_i, a) \quad (1 \leqslant i \leqslant s).$$

The latter is also a solution of (3.8), and hence

$$J_s(X;\mathbf{G}) \geqslant \sum_{1\leqslant a\leqslant X} J_s(X;\mathbf{H}') = \sum_{1\leqslant a\leqslant X} J_s(X;\mathbf{H}) = [X]J_s(X;\mathbf{H}).$$

This confirms the lower bound (3.7), and in view of our earlier discussion the proof of the theorem is complete. □

We turn now to discuss lower bounds for $J_s(X;\mathbf{F})$ for the most basic examples of reduced translation-dilation invariant systems $\mathbf{F}$.

(a) The classical system of Vinogradov. We recall the system

$$\mathbf{F} = (z^k, z^{k-1}, \dots, z)$$

of dimension 1, rank k, degree k and weight $\frac{1}{2}k(k+1)$. In this situation, the conclusion of Theorem 3.1 delivers the familiar lower bound

$$J_s(X;\mathbf{F}) \gg X^s + X^{2s-\frac{1}{2}k(k+1)}.$$

(b) The system of Parsell. We next return to the situation with $d \geqslant 2$ and the system $\mathbf{F}_d = (z_1^{i_1} z_2^{i_2} \dots z_d^{i_d} : 1 \leqslant |\mathbf{i}| \leqslant k)$. On writing

$$K_\delta = \sum_{l=1}^{k} l \binom{l+\delta-1}{l}, \tag{3.10}$$

we see that this system has weight K_d, and it follows from Theorem 3.1 that

$$\begin{aligned} J_s(X;\mathbf{F}_d) &\gg X^{sd} + X^{2sd-K_d} + \sum_{\delta=1}^{d-1} X^{d-\delta} J_s(X;\mathbf{F}_\delta) \\ &\gg X^{sd} + \sum_{\delta=1}^{d} X^{d-\delta}(X^{2s\delta-K_\delta} + X^{s\delta}). \end{aligned}$$

We therefore conclude that

$$J_s(X;\mathbf{F}_d) \gg X^{sd} + \sum_{j=1}^{d} X^{(2s-1)j+d-K_j},$$

and this establishes Theorem 1.2.

Let us consider the strength of the upper bound presented in Theorem 1.1 in the light of the lower bound just established. Define r and K as in (2.6) and (2.7). When s is large enough, we expect that $J_s(X;\mathbf{F}_d) \ll X^{2sd-K}$, and indeed this estimate is a consequence of Theorem 1.5 when $s \geqslant r(k+1)+1$. We show here that this lower bound on s cannot be relaxed substantially when k is large in terms of d.

Theorem 3.2. *Suppose that s, k and d are natural numbers, and that ν is a real number with $2d/k < \nu < 1$. Then there is a positive number $\eta = \eta(d,k)$ such that, whenever*

$$s \leqslant \frac{d}{2d+2}(1-\nu)r(k+1), \tag{3.11}$$

one has $J_{s,k,d}(X) \gg X^{2sd-K+\eta}$.

Proof. Recall the notation introduced in (3.10), and consider a natural number s satisfying (3.11). We observe that as a consequence of Theorem 1.2, one has

$$\frac{J_{s,k,d}(X)}{X^{2sd-K_d}} \gg \frac{X^{2s(d-1)+1-K_{d-1}}}{X^{2sd-K_d}} = X^{K_d-K_{d-1}+1-2s}. \tag{3.12}$$

From equations (1.2) and (1.3), one has

$$K_d = \frac{dk}{d+1}\binom{k+d}{d} = \frac{d}{d+1}(r+1)k$$

and

$$K_{d-1} = \frac{(d-1)k}{d}\binom{k+d-1}{d-1} = \frac{d-1}{k+d}(r+1)k.$$

Consequently, one finds that

$$\begin{aligned} K_d - K_{d-1} + 1 - 2s &\geqslant \frac{d}{d+1}(r+1)k - \frac{d-1}{k+d}(r+1)k + 1 \\ &\quad - \frac{d}{d+1}(1-\nu)r(k+1) \\ &\geqslant \frac{d}{d+1}(r+1)k\Big(1 - \frac{d^2-1}{d(k+d)} - (1-\nu)(1+1/k)\Big). \end{aligned}$$

Then provided that $\nu > 2d/k$, one may infer that

$$K_d - K_{d-1} + 1 - 2s > \frac{d}{d+1}(r+1)k\Big(\frac{2d-1}{k} - \frac{d}{k+d} + \frac{1}{d(k+d)}\Big) > 0.$$

One concludes therefore that there exists a positive number η such that

$$J_{s,k,d}(X) \gg X^{2sd-K_d+\eta}.$$

This completes the proof of the theorem. □

Recall that Theorem 1.1 asserts that $J_{s,k,d}(X) \ll X^{2sd-K_d+\varepsilon}$ whenever $s \geqslant r(k+1)$. Consequently, for any positive number $\nu < 1$, it follows from Theorem 3.2 that when k is large enough in terms of ν and d, such a conclusion is impossible if the constraint on s is replaced by

$$s \geqslant \frac{d}{2d+2}(1-\nu)r(k+1).$$

Thus, the conclusion of Theorem 1.1 is at worst a factor of essentially $2 + 2/d$ away from the best possible conclusion of its type.

When d is large and k is small the situation changes. Here, whenever

$$s \leqslant \frac{k}{2(k+d)}\Big(\frac{d}{d+1}\Big)(r+1)k,$$

we find that

$$K_d - K_{d-1} + 1 - 2s > \frac{d}{d+1}(r+1)k\Big(1 - \frac{d^2-1}{d(k+d)} - \frac{k}{k+d}\Big) > 0.$$

When d is large enough in terms of k, one finds that

$$\frac{k}{2(k+d)}\Big(\frac{d}{d+1}\Big)(r+1)k = r(k+1)\Big(\frac{k^2}{2(k+1)d}\Big)(1+O(k/d)).$$

Thus, when $\nu > 0$ is small enough in terms of d and k, the lower bound $J_{s,k,d}(X) \gg X^{2sd-K_d+\eta}$ holds for a positive number η provided that

$$s \leqslant \frac{(1-\nu)k^2}{2(k+1)d}r(k+1).$$

In this situation, the conclusion of Theorem 1.1 is at worst a factor of essentially $2(1+1/k)d/k$ away from the best possible conclusion of its type.

(c) The system of Arkhipov, Karatsuba and Chubarikov. Here we consider the situation with $d \geqslant 2$ and $l \geqslant 1$ in which $\mathbf{F}$ is given by (2.8). On recalling (2.10), the weight of such a system with dimension d is $K_d = \frac{1}{2}dl(l+1)^d$, whilst the corresponding weight of such a system with dimension $d-1$ is $K_{d-1} = \frac{1}{2}(d-1)l(l+1)^{d-1}$. As in the argument leading to (3.12), an application of Theorem 3.1 in this instance delivers the lower bound

$$\frac{J_s(X;\mathbf{F})}{X^{2sd-K_d}} \gg X^{K_d-K_{d-1}+1-2s}.$$

From (2.10) one finds when

$$s \leqslant \tfrac{1}{4}dl(l+1)^d(1-(1-1/d)(l+1)^{-1})$$

one has

$$K_d - K_{d-1} + 1 - 2s \geqslant 1 + \tfrac{1}{2}dl(l+1)^d(1-(1-1/d)(l+1)^{-1}) - 2s > 0.$$

Observe that

$$\tfrac{1}{4}dl(l+1)^d(1-(1-1/d)(l+1)^{-1}) = \frac{1+O(1/(dl))}{4(1+1/l)}(dl+1)((l+1)^d-1).$$

Thus, when $\nu > 0$ is small enough in terms of d and l, one finds that the lower bound $J_s(X;\mathbf{F}) \gg X^{2sd-K+\eta}$ holds for a positive number η provided that

$$s \leqslant \frac{(1-\nu)}{4(1+1/l)}(dl+1)((l+1)^d-1).$$

In this situation, the conclusion of Corollary 2.2 is at worst a factor of essentially $4(1+1/l)$ away from the best possible conclusion of its type. By way of comparison, the work of Arkhipov, Karatsuba and Chubarikov [2] would miss the best possible conclusion by a factor of order $d\log l$.

4. Preliminary manoeuvres

Our purpose in this section is to describe further notation and establish such preliminary estimates as are required to initiate the efficient congruencing procedure, with the ultimate objective of proving Theorem 2.1. With this aim in mind, let $\mathbf{F}$ be a reduced translation-dilation invariant system of polynomials having dimension d, rank r, degree k and weight K. Since the conclusion of Theorem 2.1 follows from linear algebra when $k = 1$, there is no loss of generality in supposing that $k \geqslant 2$. We consider the system $\mathbf{F}$ to be fixed throughout, and consequently suppress mention of $\mathbf{F}$ by abbreviating $J_s(X;\mathbf{F})$ to $J_s(X)$, with similar conventions in other notation as appropriate. We recall the notation introduced in (3.2) and (3.3), and note the consequence of orthogonality recorded in (3.4).

Our argument involves the investigation of systems of congruences, the singular solutions of which must be isolated for special treatment. We pause at this point to introduce a special Jacobian determinant that facilitates this treatment. First, given a polynomial $G(\mathbf{z}) \in \mathbb{Z}[z_1,\dots,z_d]$, we write $\partial_i G$ to denote the partial derivative of G with respect to the ith variable, so that

$$\partial_i G(\mathbf{z}) = \frac{\partial G}{\partial z_i}(z_1,\dots,z_d).$$

We now consider a function $\sigma : \{1,\dots,r\} \to \{1,\dots,d\}$, and define the associated Jacobian determinant $\Delta_r(\overline{\mathbf{x}};\sigma) = \Delta(\mathbf{x}_1,\dots,\mathbf{x}_r;\sigma)$ by

$$\Delta(\mathbf{x}_1,\dots,\mathbf{x}_r;\sigma) = \det\left(\partial_{\sigma(i)}F_j(\mathbf{x}_i)\right)_{1\leqslant i,j\leqslant r}. \tag{4.1}$$

We seek a choice for σ having the property that $\Delta_r(\overline{\mathbf{x}};\sigma)$ is not identically zero as a polynomial in $\overline{\mathbf{x}}$. With this goal in mind, we consider the monomials occurring in $\mathbf{F}$ and $\Delta_r(\overline{\mathbf{x}};\sigma)$, and introduce an ordering on the exponents associated with these monomials in order to ease discussion. When t is a natural number, and $\mathbf{a},\mathbf{b} \in (\mathbb{N}\cup\{0\})^t$, we say that $\mathbf{a}$ *is less than* $\mathbf{b}$ *in colex order* when there exists an index i with $1 \leqslant i \leqslant t$ such that $a_i < b_i$, and further $a_j = b_j$ for $j > i$. In such a situation, we write $\mathbf{a} \prec \mathbf{b}$, and we write $\mathbf{a} \preccurlyeq \mathbf{b}$ when $\mathbf{a} = \mathbf{b}$ or $\mathbf{a} \prec \mathbf{b}$. Next, given $\mathbf{a} \in (\mathbb{N}\cup\{0\})^t$, we write $\mathbf{x}^{\mathbf{a}}$ for the monomial $x_1^{a_1}x_2^{a_2}\dots x_t^{a_t}$. The monomials $\mathbf{x}^{\mathbf{a}}$ may now be ordered according to the colexicographical order of their indices. Notice that when $\mathbf{x}_i = (x_{i1},\dots,x_{id})$, then the monomial $\mathbf{x}_1^{\mathbf{c}_1}\dots\mathbf{x}_r^{\mathbf{c}_r}$ has smaller degree than $\mathbf{x}_1^{\mathbf{d}_1}\dots\mathbf{x}_r^{\mathbf{d}_r}$ in colex if and only if there exists an index i for which $\mathbf{c}_i \prec \mathbf{d}_i$, and further $\mathbf{c}_j = \mathbf{d}_j$ for $j > i$.

Lemma 4.1. *There exists a function $\sigma : \{1,\dots,r\} \to \{1,\dots,d\}$ such that $\Delta_r(\overline{\mathbf{x}};\sigma)$ is non-zero as a polynomial in $\mathbf{x}_1,\dots,\mathbf{x}_r$.*

Proof. We begin by interpreting the polynomials $F_j(\mathbf{z})$ in terms of the colex ordering of monomials. Recall that $k = \max_{1\leqslant j\leqslant r}\deg(F_j)$. Write

$$\mathscr{A} = \{\mathbf{a} \in (\mathbb{N}\cup\{0\})^d : 1 \leqslant a_1+\dots+a_d \leqslant k\},$$

so that the polynomials $F_j(\mathbf{z})$ are necessarily linear combinations of the monomials $\mathbf{z}^{\mathbf{a}}$ with $\mathbf{a} \in \mathscr{A}$. We put $A = \mathrm{card}(\mathscr{A})$, and label indices in such a manner that

$$\mathscr{A} = \{\mathbf{a}_1, \dots, \mathbf{a}_A\},$$

with $\mathbf{a}_1 \prec \mathbf{a}_2 \prec \dots \prec \mathbf{a}_A$ the colex ordering of the elements of $\mathscr{A}$. Thus, for $1 \leqslant j \leqslant r$, there exists an integral A-tuple $\mathbf{c}_j$, which we consider as a row vector, having the property that

$$F_j(\mathbf{z}) = \mathbf{c}_j \cdot (\mathbf{z}^{\mathbf{a}_1}, \dots, \mathbf{z}^{\mathbf{a}_A})^T.$$

Since $F_1, \dots, F_r$ are linearly independent, the matrix

$$\mathscr{C} = (\mathbf{c}_j)_{1 \leqslant j \leqslant r}$$

must have full rank, and hence there exists an invertible $r \times r$ matrix $\mathscr{M}$ with rational coefficients having the property that the product $\mathscr{R} = \mathscr{M}\mathscr{C}$ is a full rank matrix in inverted reduced row-echelon form. By the latter we mean that if $\mathscr{R}$ is the matrix

$$(\rho_{l,m})_{\substack{1 \leqslant l \leqslant r \\ 1 \leqslant m \leqslant A}},$$

then the corresponding matrix

$$(\rho_{r+1-l,A+1-m})_{\substack{1 \leqslant l \leqslant r \\ 1 \leqslant m \leqslant A}}$$

is in conventional reduced row echelon form. We define the r-tuple of polynomials $(G_1, \dots, G_r)$ by putting

$$\begin{aligned}(G_1, \dots, G_r)^T &= \mathscr{M}(F_1, \dots, F_r)^T = \mathscr{M}\mathscr{C}(\mathbf{z}^{\mathbf{a}_1}, \dots, \mathbf{z}^{\mathbf{a}_A})^T \\ &= \mathscr{R}(\mathbf{z}^{\mathbf{a}_1}, \dots, \mathbf{z}^{\mathbf{a}_A})^T. \end{aligned} \tag{4.2}$$

Let $\mathbf{z}^{\mathbf{b}_j}$ denote the leading monomial of $G_j(\mathbf{z})$ in colex, for $1 \leqslant j \leqslant r$. Since $\mathscr{R}$ is in inverted reduced row echelon form, we have

$$\mathbf{b}_1 \prec \mathbf{b}_2 \prec \dots \prec \mathbf{b}_r. \tag{4.3}$$

We now define the function $\sigma : \{1, \dots, r\} \to \{1, \dots, d\}$ as follows. When $1 \leqslant i \leqslant r$, we take $\sigma(i)$ to be the smallest index h with the property that $b_{ih} > 0$.

It remains to verify that with the choice of the function σ just made, the polynomial $\Delta_r(\overline{\mathbf{x}}; \sigma)$ is non-zero. Let $\mathbf{e}_i$ denote the d-dimensional vector whose ith coordinate is equal to 1, all of whose remaining coordinates are 0. When $1 \leqslant n \leqslant r$, define the Jacobian determinant $D_n(\overline{\mathbf{x}}; \sigma) = D(\mathbf{x}_1, \dots, \mathbf{x}_n; \sigma)$ by

$$D(\mathbf{x}_1, \dots, \mathbf{x}_n; \sigma) = \det(\partial_{\sigma(i)} G_j(\mathbf{x}_i))_{1 \leqslant i,j \leqslant n}.$$

We proceed by induction to show that for $1 \leqslant n \leqslant r$, the determinant $D_n(\overline{\mathbf{x}}; \sigma)$ may be expanded in the shape

$$D_n(\overline{\mathbf{x}}; \sigma) = D_n^{(1)}(\overline{\mathbf{x}}; \sigma) + D_n^{(2)}(\overline{\mathbf{x}}; \sigma), \tag{4.4}$$

where

$$D_n^{(1)}(\overline{\mathbf{x}}; \sigma) = \prod_{j=1}^{n} b_{j,\sigma(j)} \mathbf{x}_j^{\mathbf{b}_j - \mathbf{e}_{\sigma(j)}}, \tag{4.5}$$

and $D_n^{(2)}(\overline{\mathbf{x}};\sigma)$ is of smaller degree in colex than $D_n^{(1)}(\overline{\mathbf{x}};\sigma)$. Notice that the definition of σ implies that $b_{1,\sigma(1)}\dots b_{n,\sigma(n)}\neq 0$. Then having established the inductive hypothesis for $n=r$, it follows that $D_r(\overline{\mathbf{x}};\sigma)$ contains the monomial

$$\prod_{j=1}^{r}\mathbf{x}_j^{\mathbf{b}_j-\mathbf{e}_{\sigma(j)}}$$

with a non-zero coefficient, and hence $D_r(\overline{\mathbf{x}};\sigma)$ must be a non-zero polynomial. In this way, the proof of the inductive hypothesis will facilitate the proof of the lemma.

When $n=1$, we have $D_n(\overline{\mathbf{x}};\sigma)=\partial_{\sigma(1)}G_1(\mathbf{x}_1)$, and hence it follows at once that (4.4) holds with

$$D_1^{(1)}(\overline{\mathbf{x}};\sigma)=b_{1,\sigma(1)}\mathbf{x}_1^{\mathbf{b}_1-\mathbf{e}_{\sigma(1)}},$$

for some polynomial $D_1^{(2)}(\overline{\mathbf{x}};\sigma)$ of degree smaller in colex than $\mathbf{x}_1^{\mathbf{b}_1-\mathbf{e}_{\sigma(1)}}$. Thus the inductive hypothesis holds with $n=1$.

Suppose next that the inductive hypothesis has been established already for $1\leqslant n<u$. When $1\leqslant i,j\leqslant u$, define the polynomials θ_{ij} by putting

$$\theta_{ij}=\begin{cases} b_{u,\sigma(u)}\mathbf{x}_u^{\mathbf{b}_u-\mathbf{e}_{\sigma(u)}}, & \text{when } i=j=u,\\ 0, & \text{when } (i,j)\neq(u,u).\end{cases}$$

Then the determinant $D_u(\overline{\mathbf{x}};\sigma)$ has the expansion

$$D_u(\overline{\mathbf{x}};\sigma)=\theta_{uu}D_{u-1}(\overline{\mathbf{x}};\sigma)+\det(\partial_{\sigma(i)}G_j(\mathbf{x}_i)-\theta_{ij})_{1\leqslant i,j\leqslant u}.$$

On making use of the inductive hypothesis with $n=u-1$ in order to expand $D_{u-1}(\overline{\mathbf{x}};\sigma)$, we deduce that

$$\begin{aligned} D_u(\overline{\mathbf{x}};\sigma)&=\theta_{uu}(D_{u-1}^{(1)}(\overline{\mathbf{x}};\sigma)+D_{u-1}^{(2)}(\overline{\mathbf{x}};\sigma))+\det(\partial_{\sigma(i)}G_j(\mathbf{x}_i)-\theta_{ij})_{1\leqslant i,j\leqslant u}\\ &=D_u^{(1)}(\overline{\mathbf{x}};\sigma)+D_u^{(0)}(\overline{\mathbf{x}};\sigma),\end{aligned}\tag{4.6}$$

where

$$D_u^{(0)}(\overline{\mathbf{x}};\sigma)=b_{u,\sigma(u)}\mathbf{x}_u^{\mathbf{b}_u-\mathbf{e}_{\sigma(u)}}D_{u-1}^{(2)}(\overline{\mathbf{x}};\sigma)+\det(\partial_{\sigma(i)}G_j(\mathbf{x}_i)-\theta_{ij})_{1\leqslant i,j\leqslant u}.\tag{4.7}$$

Every term on the right hand side of (4.7) may be expanded as a linear combination of terms of the shape

$$\mathbf{x}_1^{\mathbf{h}_1-\mathbf{e}_{\sigma(1)}}\dots\mathbf{x}_u^{\mathbf{h}_u-\mathbf{e}_{\sigma(u)}},\tag{4.8}$$

in which for some permutation $\pi:\{1,\dots,u\}\to\{1,\dots,u\}$, there exist d-tuples $\mathbf{b}_1',\dots,\mathbf{b}_u'$ with $(\mathbf{b}_1',\dots,\mathbf{b}_u')\preccurlyeq(\mathbf{b}_1,\dots,\mathbf{b}_u)$ and $\mathbf{b}_l'\preccurlyeq\mathbf{b}_l$ $(1\leqslant l\leqslant u)$, and having the property that

$$(\mathbf{h}_1,\dots,\mathbf{h}_u)=(\mathbf{b}_{\pi(1)}',\dots,\mathbf{b}_{\pi(u)}').$$

Moreover, in the event that π is the identity permutation, then one has $(\mathbf{b}_1',\dots,\mathbf{b}_u')\prec(\mathbf{b}_1,\dots,\mathbf{b}_u)$.

If π is the identity permutation, then

$$(\mathbf{h}_1,\dots,\mathbf{h}_u)=(\mathbf{b}_1',\dots,\mathbf{b}_u')\prec(\mathbf{b}_1,\dots,\mathbf{b}_u).$$

Since the (strict) colex ordering between tuples is not reversed on componentwise addition, a modicum of thought confirms that

$$(\mathbf{h}_1 - \mathbf{e}_{\sigma(1)}, \dots, \mathbf{h}_u - \mathbf{e}_{\sigma(u)}) \prec (\mathbf{b}_1 - \mathbf{e}_{\sigma(1)}, \dots, \mathbf{b}_u - \mathbf{e}_{\sigma(u)}). \tag{4.9}$$

Unfortunately, our notation is somewhat opaque, and so it may be worthwhile to spell out the details underlying this deduction. Since $(\mathbf{h}_1, \dots, \mathbf{h}_u) \prec (\mathbf{b}_1, \dots, \mathbf{b}_u)$, there exists an index l with the property that $\mathbf{h}_l \prec \mathbf{b}_l$, and further $\mathbf{h}_j = \mathbf{b}_j$ for $j > l$. But then there exists an index $m = m(l)$ with the property that $h_{lm} < b_{lm}$, and further $h_{lj} = b_{lj}$ for $j > m$. On recalling that the definition of the function σ implies that $b_{lj} = 0$ for $j < \sigma(l)$, we deduce that $m(l) \geqslant \sigma(l)$. Thus we find that $\mathbf{h}_l - \mathbf{e}_{\sigma(l)} \prec \mathbf{b}_l - \mathbf{e}_{\sigma(l)}$, and further that $\mathbf{h}_j - \mathbf{e}_{\sigma(j)} = \mathbf{b}_j - \mathbf{e}_{\sigma(j)}$ for $j > l$. We therefore conclude that (4.9) holds, as we had previously asserted.

Suppose next that π is not the identity permutation, and let i be maximal with $\pi(i) \neq i$. Then for $j > i$ one has $j = \pi(j)$, and hence

$$\mathbf{h}_j = \mathbf{b}'_{\pi(j)} = \mathbf{b}'_j \preccurlyeq \mathbf{b}_j.$$

Since $j = \pi(j)$ for $j > i$, it follows that $i > \pi(i)$, and hence from (4.3) we see that

$$\mathbf{h}_i = \mathbf{b}'_{\pi(i)} \preccurlyeq \mathbf{b}_{\pi(i)} \prec \mathbf{b}_i.$$

It follows that $(\mathbf{h}_1, \dots, \mathbf{h}_u) \prec (\mathbf{b}_1, \dots, \mathbf{b}_u)$, so that as above one finds that the relation (4.9) holds also in the situation that π is not the identity permutation. In view of (4.8), it follows that $D_u^{(0)}(\overline{\mathbf{x}}; \sigma)$ is of smaller degree in colex than the polynomial $D_u^{(1)}(\overline{\mathbf{x}}; \sigma)$ defined by means of (4.5). We therefore deduce from (4.6) that the relation (4.4) holds with $n = u$, and with $D_u^{(2)}(\overline{\mathbf{x}}; \sigma)$ of smaller degree in colex than $D_u^{(1)}(\overline{\mathbf{x}}; \sigma)$. We have consequently established the inductive hypothesis with $n = u$, and hence the inductive hypothesis holds for $1 \leqslant n \leqslant r$. In particular, the determinant $D_r(\overline{\mathbf{x}}; \sigma)$ is a non-zero polynomial.

We now reverse course in order to relate the non-vanishing of $D_r(\overline{\mathbf{x}}; \sigma)$ to the non-vanishing of $\Delta_r(\overline{\mathbf{x}}; \sigma)$. For each l with $1 \leqslant l \leqslant d$, it follows from (4.2) that one has the identity

$$(\partial_l G_1, \dots, \partial_l G_r)^T = \mathscr{M}(\partial_l F_1, \dots, \partial_l F_r)^T,$$

whence

$$(\partial_{\sigma(i)} G_j(\mathbf{x}_i))_{1 \leqslant i,j \leqslant r}^T = \mathscr{M}(\partial_{\sigma(i)} F_j(\mathbf{x}_i))_{1 \leqslant i,j \leqslant r}^T.$$

Consequently, one has

$$D_r(\overline{\mathbf{x}}; \sigma) = (\det \mathscr{M}) \Delta_r(\overline{\mathbf{x}}; \sigma).$$

But the matrix $\mathscr{M}$ is invertible, so that $\det \mathscr{M} \neq 0$. Our conclusion that $D_r(\overline{\mathbf{x}}; \sigma)$ is a non-zero polynomial therefore forces us to conclude that $\Delta_r(\overline{\mathbf{x}}; \sigma)$ is also a non-zero polynomial. This completes the proof of the lemma. □

Henceforth, we fix our choice of the function $\sigma : \{1, \dots, r\} \to \{1, \dots, d\}$ so that $\Delta_r(\overline{\mathbf{x}}; \sigma)$ is a non-zero polynomial, as permitted by the conclusion of Lemma 4.1, and we write $\Delta_r(\overline{\mathbf{x}})$ for $\Delta_r(\overline{\mathbf{x}}; \sigma)$.

We next turn to the task of introducing notation with which to describe the mean values central to our methods, as well as exponents with which to bound these mean values. We refer to the exponent $\lambda_s = \lambda_s(\mathbf{F})$ as *permissible* when, for each positive number ε, and for any real number X sufficiently large in terms of s, $\mathbf{F}$ and ε, one has $J_s(X;\mathbf{F}) \ll X^{\lambda_s+\varepsilon}$. Define λ_s^* to be the infimum of the set of exponents λ_s permissible for s and $\mathbf{F}$, and then put $\eta_s = \lambda_s^* - 2sd + K$. Thus, whenever X is sufficiently large in terms of s, $\mathbf{F}$ and ε, one has

$$J_s(X) \ll X^{\lambda_s^*+\varepsilon}, \tag{4.10}$$

where

$$\lambda_s^* = 2sd - K + \eta_s. \tag{4.11}$$

Note that, in view of the lower bound supplied by Theorem 3.1 and the trivial estimate $J_s(X) \leqslant X^{2sd}$, one has $0 \leqslant \eta_s \leqslant K$ for $s \in \mathbb{N}$.

We take δ to be a small positive number to be fixed shortly. Let u be a natural number with $u \geqslant k$, put $s_0 = ur$, and fix a natural number s with $s \geqslant s_0$. Our goal is to show that $\lambda_{s+r}^* = 2(s+r)d - K$, whence $\eta_{s+r} = 0$. From this and the definition of λ_{s+r}^*, it follows that there exists a sequence of natural numbers $(X_n)_{n=1}^{\infty}$, tending to infinity, with the property that

$$J_{s+r}(X_n) > X_n^{\lambda_{s+r}^*-\delta} \quad (n \in \mathbb{N}). \tag{4.12}$$

In view of (4.10), when X_n is sufficiently large and $X_n^{\delta^2} < Y \leqslant X_n$, we also have the corresponding upper bound

$$J_{s+r}(Y) < Y^{\lambda_{s+r}^*+\delta}. \tag{4.13}$$

Notice that since $s \geqslant s_0$, the trivial inequality $|f(\boldsymbol{\alpha};X)| \leqslant X^d$ leads to the upper bound

$$J_{s+r}(X) \leqslant X^{2(s-s_0)d} \oint |f(\boldsymbol{\alpha};X)|^{2s_0+2r}\,\mathrm{d}\boldsymbol{\alpha} = X^{2(s-s_0)d} J_{s_0+r}(X).$$

It follows that one has $\eta_{s+r} \leqslant \eta_{s_0+r}$, and so we are free to restrict attention to the special case $s = s_0$. Finally, we take N to be a natural number sufficiently large in terms of s and $\mathbf{F}$. We then put

$$\theta = N^{-1/2}(r/s)^{N+2}, \tag{4.14}$$

and fix δ to be a positive number with $\delta < (Ns)^{-3N}$, so that δ is small compared to θ. We now consider a fixed element $X = X_n$ of the sequence (X_n), which we may assume to be sufficiently large in terms of s, $\mathbf{F}$, N and δ, and we put $M = X^{\theta}$. Thus, in particular, one has $X^{\delta} < M^{1/N}$.

Let p be a fixed prime number with $M < p \leqslant 2M$ to be chosen in due course. That such a prime exists is a consequence of the Prime Number Theorem. Recall the notation introduced in (3.2). When c is a non-negative integer, and $\boldsymbol{\xi} \in \mathbb{Z}^d$ and $\boldsymbol{\alpha} \in [0,1)^r$, define

$$\mathfrak{f}_c(\boldsymbol{\alpha};\boldsymbol{\xi}) = \sum_{\substack{1 \leqslant \mathbf{x} \leqslant X \\ \mathbf{x} \equiv \boldsymbol{\xi} \pmod{p^c}}} e(\psi(\mathbf{x};\boldsymbol{\alpha})). \tag{4.15}$$

We need to equip ourselves with exponential sums of a type related to that defined in (4.15), but possessing inherently non-singular structure. We introduce the notion of *well-conditioned* r-tuples $(\mathbf{x}_1, \dots, \mathbf{x}_r)$ with $\mathbf{x}_i \in \mathbb{Z}^d$ $(1 \leqslant i \leqslant r)$. We denote by $\Xi_c(\boldsymbol{\xi})$ the set of r-tuples $\overline{\boldsymbol{\xi}} = (\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_r) \in (\mathbb{Z}^d)^r$, with

$$1 \leqslant \boldsymbol{\xi}_i \leqslant p^{c+1} \quad \text{and} \quad \boldsymbol{\xi}_i \equiv \boldsymbol{\xi} \pmod{p^c} \quad (1 \leqslant i \leqslant r),$$

and satisfying the property that $\Delta(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_r) \not\equiv 0 \pmod{p^{(K-r)c+1}}$. In addition, write $\Sigma_r = \{1, -1\}^r$, and consider an element $\boldsymbol{\sigma}$ of Σ_r. Recalling the definition (4.15), we then put

$$\mathfrak{F}_c^{\boldsymbol{\sigma}}(\boldsymbol{\alpha}; \boldsymbol{\xi}) = \sum_{\overline{\boldsymbol{\xi}} \in \Xi_c(\boldsymbol{\xi})} \prod_{i=1}^{r} \mathfrak{f}_{c+1}(\sigma_i \boldsymbol{\alpha}; \boldsymbol{\xi}_i). \tag{4.16}$$

Next we introduce the two mean values that underly our arguments. When a and b are non-negative integers, and $\boldsymbol{\sigma}, \boldsymbol{\tau} \in \Sigma_r$, we define

$$I_{a,b}^{\boldsymbol{\sigma}}(X; \boldsymbol{\xi}, \boldsymbol{\eta}) = \oint |\mathfrak{F}_a^{\boldsymbol{\sigma}}(\boldsymbol{\alpha}; \boldsymbol{\xi})^2 \mathfrak{f}_b(\boldsymbol{\alpha}; \boldsymbol{\eta})^{2s}| \, \mathrm{d}\boldsymbol{\alpha}, \tag{4.17}$$

and

$$K_{a,b}^{\boldsymbol{\sigma},\boldsymbol{\tau}}(X; \boldsymbol{\xi}, \boldsymbol{\eta}) = \oint |\mathfrak{F}_a^{\boldsymbol{\sigma}}(\boldsymbol{\alpha}; \boldsymbol{\xi})^2 \mathfrak{F}_b^{\boldsymbol{\tau}}(\boldsymbol{\alpha}; \boldsymbol{\eta})^{2u}| \, \mathrm{d}\boldsymbol{\alpha}. \tag{4.18}$$

We then define

$$I_{a,b}(X) = \max_{1 \leqslant \boldsymbol{\xi} \leqslant p^a} \max_{1 \leqslant \boldsymbol{\eta} \leqslant p^b} \max_{\boldsymbol{\sigma} \in \Sigma_r} I_{a,b}^{\boldsymbol{\sigma}}(X; \boldsymbol{\xi}, \boldsymbol{\eta}), \tag{4.19}$$

and

$$K_{a,b}(X) = \max_{1 \leqslant \boldsymbol{\xi} \leqslant p^a} \max_{1 \leqslant \boldsymbol{\eta} \leqslant p^b} \max_{\boldsymbol{\sigma}, \boldsymbol{\tau} \in \Sigma_r} K_{a,b}^{\boldsymbol{\sigma},\boldsymbol{\tau}}(X; \boldsymbol{\xi}, \boldsymbol{\eta}). \tag{4.20}$$

We stress that, although these mean values depend on our choice of p, this dependence will shortly be rendered irrelevant when we fix our choice of p once and for all. Consequently, we suppress mention of p in our notation.

Finally, following the simplifying notational device of [21, §3], we define the normalised magnitude of the mean values $J_{s+r}(X)$ and $K_{a,b}(X)$ as follows. We define $[\![J_{s+r}(X)]\!]$ by means of the relation

$$J_{s+r}(X) = X^{2(s+r)d-K}[\![J_{s+r}(X)]\!], \tag{4.21}$$

and when $0 \leqslant a < b$, we define $[\![K_{a,b}(X)]\!]$ by means of the relation

$$K_{a,b}(X) = (X/M^a)^{2rd-K}(X/M^b)^{2sd}[\![K_{a,b}(X)]\!]. \tag{4.22}$$

Note that in view of (4.11), the lower bound (4.12) implies that

$$[\![J_{s+r}(X)]\!] > X^{\eta_{s+r}-\delta}, \tag{4.23}$$

while the upper bound (4.13) ensures that whenever $X^{\delta^2} < Y \leqslant X$, then

$$[\![J_{s+r}(Y)]\!] < Y^{\eta_{s+r}+\delta}. \tag{4.24}$$

5. Auxiliary mean values

We collect together in this section several mean value estimates that facilitate our subsequent analysis. We begin by exploiting the translation-dilation invariance of the system $\mathbf{F}$ so as to bound an analogue of $J_s(X)$ in which variables are restricted to an arithmetic progression. This argument will be familiar to those who work on such problems.

Lemma 5.1. *Suppose that c is a non-negative integer with $c\theta \leqslant 1$. Then for each natural number t, one has*

$$\max_{1\leqslant \boldsymbol{\xi}\leqslant p^c} \oint |\mathfrak{f}_c(\boldsymbol{\alpha};\boldsymbol{\xi})|^{2t}\,\mathrm{d}\boldsymbol{\alpha} \ll_t J_t(X/M^c). \tag{5.1}$$

Proof. Let $\boldsymbol{\xi}$ be an integral d-tuple with $1 \leqslant \boldsymbol{\xi} \leqslant p^c$. By orthogonality, it follows from (4.15) that the integral on the left hand side of (5.1) is bounded above by the number of integral solutions of the system

$$\sum_{i=1}^{t}(\mathbf{F}(p^c\mathbf{y}_i+\boldsymbol{\xi})-\mathbf{F}(p^c\mathbf{z}_i+\boldsymbol{\xi}))=\mathbf{0}, \tag{5.2}$$

with $0 \leqslant \overline{\mathbf{y}},\overline{\mathbf{z}} \leqslant X/p^c$. The translation-dilation invariance of the system $\mathbf{F}$ discussed in the context of (2.1) and (2.4) shows that $\overline{\mathbf{y}},\overline{\mathbf{z}}$ is an integral solution of (5.2) if and only if

$$\sum_{i=1}^{t}(\mathbf{F}(\mathbf{y}_i)-\mathbf{F}(\mathbf{z}_i))=\mathbf{0}. \tag{5.3}$$

But on recalling (3.3) and employing orthogonality again, we perceive that the number of integral solutions of (5.3) with $0 \leqslant \overline{\mathbf{y}},\overline{\mathbf{z}} \leqslant X/p^c$ is equal to

$$\oint |1+f(\boldsymbol{\alpha};X/p^c)|^{2t}\,\mathrm{d}\boldsymbol{\alpha} \ll_t 1+\oint |f(\boldsymbol{\alpha};X/p^c)|^{2t}\,\mathrm{d}\boldsymbol{\alpha}.$$

Thus we obtain the upper bound

$$\oint |\mathfrak{f}_c(\boldsymbol{\alpha};\boldsymbol{\xi})|^{2t}\,\mathrm{d}\boldsymbol{\alpha} \ll 1+J_t(X/p^c).$$

Since the condition $c\theta \leqslant 1$ ensures that $X/M^c \geqslant 1$, a consideration of diagonal solutions ensures that $J_t(X/M^c) \geqslant 1$, and the conclusion of the lemma follows on noting that $J_t(X/p^c) \leqslant J_t(X/M^c)$. □

Singular solutions associated with the vanishing of $\Delta_r(\overline{\mathbf{x}})$ are difficult to control, and so we prepare a lemma to bound their number. We first introduce some additional notation. Let $\mathbb{F}$ be a field. We restrict attention to either the field of rational numbers $\mathbb{Q}$, or else the finite field of p elements $\mathbb{F}_p$. The coefficients of $\Delta_r(\overline{\mathbf{x}})$ embed into both of these fields. Suppose that $\mathscr{A} \subseteq \mathbb{F}$ is finite, and write $A=\mathrm{card}(\mathscr{A})$. Let t be a natural number. We denote by $\mathscr{S}_t(\mathscr{A};\mathbb{F})$ the set of t-tuples $(\mathbf{x}_1,\dots,\mathbf{x}_t) \in (\mathscr{A}^d)^t$ having the property that the determinants $\Delta_r(\mathbf{x}_{j_1},\dots,\mathbf{x}_{j_r})$ vanish for all r-tuples $(j_1,\dots,j_r)$ with $1 \leqslant \mathbf{j} \leqslant t$. Before announcing our main estimate for $\mathrm{card}(\mathscr{S}_t(\mathscr{A};\mathbb{F}))$, we first recall a

familiar lemma bounding the number of zeros of polynomials in many variables (see [18, Lemma 2], for example).

Lemma 5.2. *Let $\Upsilon \in \mathbb{F}[y_1, \dots, y_u]$ be a non-trivial polynomial of total degree κ. Then the number of solutions of the equation $\Upsilon(y_1, \dots, y_u) = 0$ with $\mathbf{y} \in \mathscr{A}^u$ is at most κA^{u-1}.*

Proof. We proceed inductively. When $u = 1$, the desired conclusion is a consequence of Lagrange's theorem. Suppose then that the conclusion of the lemma has been established for $1 \leqslant u < v$, and let $\Psi \in \mathbb{F}[y_1, \dots, y_v]$ be a non-trivial polynomial of total degree κ. By relabelling variables if necessary, we may suppose that Ψ is explicit in y_v. Let the degree of Ψ with respect to y_v be ω, and let the coefficient of y_v^ω be $\Phi(y_1, \dots, y_{v-1})$. Then Φ is a non-trivial polynomial in $v-1$ variables of degree at most $\kappa - \omega$. By the inductive hypothesis, the number of solutions of $\Phi(y_1, \dots, y_{v-1}) = 0$ with $(y_1, \dots, y_{v-1}) \in \mathscr{A}^{v-1}$ is at most $(\kappa - \omega)A^{v-2}$. Then the number of v-tuples $(y_1, \dots, y_v) \in \mathscr{A}^v$ satisfying $\Phi(y_1, \dots, y_{v-1}) = 0$ is at most $(\kappa-\omega)A^{v-1}$. Meanwhile, when $\Phi(y_1, \dots, y_{v-1}) \neq 0$ and $\Psi(y_1, \dots, y_v) = 0$, then y_v satisfies a non-trivial polynomial of degree ω determined by $y_1, \dots, y_{v-1}$. So there are at most ωA^{v-1} solutions of $\Psi(\mathbf{y}) = 0$ with $\mathbf{y} \in \mathscr{A}^v$ and $\Phi(y_1, \dots, y_{v-1}) \neq 0$. We conclude that the total number of solutions of $\Psi(\mathbf{y}) = 0$ with $\mathbf{y} \in \mathscr{A}^v$ is at most $(\kappa - \omega)A^{v-1} + \omega A^{v-1} = \kappa A^{v-1}$, and hence the inductive hypothesis follows for $u = v$. The desired conclusion therefore follows by induction. □

Lemma 5.3. *Suppose that $\Delta_r(\overline{\mathbf{x}})$ is not identically zero as a polynomial in $\mathbb{F}$. Then*

$$\mathrm{card}(\mathscr{S}_t(\mathscr{A}; \mathbb{F})) \ll A^{t(d-1)+r-1}.$$

Proof. The conclusion of the lemma is trivial when $t < r$, so we may assume that $t \geqslant r$. We define a sequence of non-zero polynomials $\mathscr{D}_i(\overline{\mathbf{x}}) = \mathscr{D}_i(\mathbf{x}_1, \dots, \mathbf{x}_i)$ $(0 \leqslant i \leqslant r)$ as follows. We begin by setting $\mathscr{D}_r(\overline{\mathbf{x}}) = \Delta_r(\overline{\mathbf{x}})$. Suppose then that for some $l \geqslant 1$ we have constructed the polynomials $\mathscr{D}_i(\overline{\mathbf{x}})$ for $l \leqslant i \leqslant r$. Amongst the monomials $\mathbf{x}_1^{\mathbf{h}_1} \dots \mathbf{x}_l^{\mathbf{h}_l}$ occurring in $\mathscr{D}_l(\overline{\mathbf{x}})$, let $\mathbf{b}_l$ denote the largest of the d-tuples $\mathbf{h}_l$ in colex. It follows that there exist polynomials $\mathscr{D}_{l-1}(\overline{\mathbf{x}})$ and $\mathscr{R}_l(\overline{\mathbf{x}})$ having the property that

$$\mathscr{D}_l(\mathbf{x}_1, \dots, \mathbf{x}_l) = \mathscr{D}_{l-1}(\mathbf{x}_1, \dots, \mathbf{x}_{l-1})\mathbf{x}_l^{\mathbf{b}_l} + \mathscr{R}_l(\mathbf{x}_1, \dots, \mathbf{x}_l). \tag{5.4}$$

We may suppose that $\mathscr{D}_{l-1}(\overline{\mathbf{x}})$ is non-zero, and that every monomial $\mathbf{x}_1^{\mathbf{h}_1} \dots \mathbf{x}_l^{\mathbf{h}_l}$ occurring in $\mathscr{R}_l(\overline{\mathbf{x}})$ satisfies $\mathbf{h}_l \prec \mathbf{b}_l$. In this way, we have defined polynomials $\mathscr{D}_i(\overline{\mathbf{x}})$ for $0 \leqslant i \leqslant r$. Notice here that $\mathscr{D}_0(\overline{\mathbf{x}})$ is a non-zero element of $\mathbb{F}$.

Consider an integer j with $1 \leqslant j \leqslant r$, and denote by $\mathscr{B}_j$ the set of all j-element subsets of $\{1, 2, \dots, t\}$. We define $\mathscr{T}_j$ to be the set of t-tuples $(\mathbf{x}_1, \dots, \mathbf{x}_t) \in (\mathscr{A}^d)^t$ satisfying the property that (a) for each subset $\{l_1, \dots, l_j\}$ in $\mathscr{B}_j$, one has $\mathscr{D}_j(\mathbf{x}_{l_1}, \dots, \mathbf{x}_{l_j}) = 0$, and (b) whenever $i < j$, there exists a subset $\{m_1, \dots, m_i\}$ in $\mathscr{B}_i$ such that $\mathscr{D}_i(\mathbf{x}_{m_1}, \dots, \mathbf{x}_{m_i}) \neq 0$. A moment of reflection reveals that $\mathscr{S}_t(\mathscr{A}; \mathbb{F})$ is the union of $\mathscr{T}_1, \mathscr{T}_2, \dots, \mathscr{T}_r$.

We next seek to bound $\mathrm{card}(\mathscr{T}_j)$ for each integer j with $1 \leqslant j \leqslant r$. Consider a t-tuple $(\mathbf{x}_1, \dots, \mathbf{x}_t) \in \mathscr{T}_j$. There exists a subset $\{m_1, \dots, m_{j-1}\}$ in $\mathscr{B}_{j-1}$

with $\mathscr{D}_{j-1}(\mathbf{x}_{m_1},\dots,\mathbf{x}_{m_{j-1}}) \neq 0$ having the property that for each index m with $1 \leqslant m \leqslant t$ for which $m \notin \{m_1,\dots,m_{j-1}\}$, one has $\mathscr{D}_j(\mathbf{x}_{m_1},\dots,\mathbf{x}_{m_{j-1}},\mathbf{x}_m) = 0$. In view of the relation (5.4), the latter equation implies that the d-tuple $\mathbf{x}_m$ satisfies

$$\mathscr{D}_{j-1}(\mathbf{x}_{m_1},\dots,\mathbf{x}_{m_{j-1}})\mathbf{x}_m^{\mathbf{b}_j} + \mathscr{R}_j(\mathbf{x}_{m_1},\dots,\mathbf{x}_{m_{j-1}},\mathbf{x}_m) = 0, \tag{5.5}$$

in which the second term on the left hand side is of smaller degree in $\mathbf{x}_m$ in colex than the first term. Notice that since $\mathscr{D}_{j-1}(\mathbf{x}_{m_1},\dots,\mathbf{x}_{m_{j-1}}) \neq 0$, then in particular the equation (5.5) is non-trivial as a polynomial equation in $\mathbf{x}_m$. In this way, we deduce from Lemma 5.2 that for each index m with $1 \leqslant m \leqslant t$ for which $m \notin \{m_1,\dots,m_{j-1}\}$, the number of d-tuples $\mathbf{x}_m \in \mathscr{A}^d$ satisfying (5.5) is at most

$$(b_{j1} + \dots + b_{jd})A^{d-1} \leqslant (k-1)A^{d-1}.$$

The total number of choices of $\mathbf{x}_m \in \mathscr{A}^d$ for all such indices m is therefore at most $((k-1)A^{d-1})^{t-(j-1)}$. The number of choices for $(\mathbf{x}_{m_1},\dots,\mathbf{x}_{m_{j-1}}) \in (\mathscr{A}^d)^{j-1}$, meanwhile, is at most $A^{d(j-1)}$. Since a trivial estimate confirms that $\mathrm{card}(\mathscr{B}_{j-1}) \leqslant t^{j-1}$, we deduce that

$$\mathrm{card}(\mathscr{T}_j) \leqslant t^{j-1}((k-1)A^{d-1})^{t-j+1}(A^d)^{j-1} \ll A^{t(d-1)+j-1}.$$

Combining the contributions of $\mathscr{T}_1,\dots,\mathscr{T}_r$, therefore, we conclude that

$$\mathrm{card}(\mathscr{S}_t(\mathscr{A};\mathbb{F})) = \sum_{j=1}^{r} \mathrm{card}(\mathscr{T}_j) \ll \sum_{j=1}^{r} A^{t(d-1)+j-1} \ll A^{t(d-1)+r-1}.$$

This completes the proof of the lemma. □

We are now equipped to initiate the iterative procedure. It is at this point that we fix our choice for the prime number p.

Lemma 5.4. *There exists a prime number p with $M < p \leqslant 2M$ for which $J_{s+r}(X) \ll M^{2sd} I_{0,1}(X)$.*

Proof. The mean value $J_{s+r}(X)$ counts the number of integral solutions of the system

$$\sum_{i=1}^{2(s+r)} (-1)^i \mathbf{F}(\mathbf{x}_i) = \mathbf{0}, \tag{5.6}$$

with $1 \leqslant \overline{\mathbf{x}} \leqslant X$. Let T_0 denote the number of such solutions in which

$$\Delta(\mathbf{x}_{i_1},\dots,\mathbf{x}_{i_r}) = 0 \tag{5.7}$$

for all indices i_l $(1 \leqslant l \leqslant r)$ satisfying

$$1 \leqslant \mathbf{i} \leqslant 2(s+r). \tag{5.8}$$

Also, let T_1 denote the corresponding number of solutions in which (5.7) fails to hold for some index $\mathbf{i}$ satisfying (5.8). Then $J_{s+r}(X) = T_0 + T_1$.

We first consider T_0. Put $\mathscr{A} = \{1,2,\dots,[X]\}$ and $\mathscr{S} = \mathscr{S}_{2s+2r}(\mathscr{A};\mathbb{Q})$. Then on recalling the discussion in the preamble to the statement of Lemma

5.2, we see that whenever $\mathbf{x}$ is counted by T_0, one has $\mathbf{x} \in \mathscr{S}$. It therefore follows from Lemma 5.3 that

$$T_0 \leqslant \mathrm{card}(\mathscr{S}_{2s+2r}(\mathscr{A};\mathbb{Q})) \ll X^{2(s+r)(d-1)+r-1}.$$

Note that our hypotheses ensure that $s \geqslant rk$. Then since

$$K = \sum_{j=1}^{r} k_j \leqslant rk,$$

we find that $s \geqslant K$. In view of the lower bound on $J_{s+r}(X)$ available from Theorem 3.1, we deduce that

$$T_0 \ll X^{2(s+r)d-2s-1} \ll X^{2(s+r)d-K-1} \ll X^{-1}J_{s+r}(X). \tag{5.9}$$

We next turn to the solutions counted by T_1. We begin by examining the determinant $\Delta_r(\overline{\mathbf{z}})$. On recalling (4.1), we see that the (j,i)-th entry of the matrix associated with the determinant $\Delta_r(\overline{\mathbf{z}})$ has degree at most $k_j - 1$. As a polynomial in $\overline{\mathbf{z}}$, therefore, we see that the degree of $\Delta_r(\overline{\mathbf{z}})$ satisfies

$$\deg \Delta_r(\overline{\mathbf{z}}) \leqslant \sum_{j=1}^{r}(k_j - 1) = K - r.$$

The coefficients of the monomial entries of $\Delta_r(\overline{\mathbf{z}})$ depend at most on $\mathbf{F}$, and so for sufficiently large values of X, the sum of the absolute values of these coefficients is bounded above by X. Thus we see that

$$\max_{1 \leqslant \overline{\mathbf{z}} \leqslant X} |\Delta_r(\overline{\mathbf{z}})| \leqslant X^K.$$

Let $\mathscr{P}$ denote any set of $[K/\theta]+1$ distinct prime numbers with $M < p \leqslant 2M$. Such a set exists by the Prime Number Theorem, since we are at liberty to assume X to be large enough in terms of K and θ. It follows that

$$\prod_{p \in \mathscr{P}} p > M^{K/\theta} = X^K \geqslant \max_{1 \leqslant \overline{\mathbf{z}} \leqslant X} |\Delta_r(\overline{\mathbf{z}})|.$$

Consequently, whenever $1 \leqslant \overline{\mathbf{z}} \leqslant X$ and $\Delta_r(\overline{\mathbf{z}}) \neq 0$, then there exists a prime $p \in \mathscr{P}$ for which $p \nmid \Delta_r(\overline{\mathbf{z}})$. In particular, for each solution $\overline{\mathbf{x}}$ of (5.6) counted by T_1, there exists an index $\mathbf{i}$ satisfying (5.8) and a prime $p \in \mathscr{P}$ for which $\Delta_r(\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_r}) \not\equiv 0 \pmod{p}$.

Let $\mathscr{I}$ denote the set of all indices $\mathbf{i}$ satisfying (5.8), and define $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\mathbf{i})$ by putting $\boldsymbol{\sigma}(\mathbf{i}) = ((-1)^{i_1}, \dots, (-1)^{i_r})$. Also, put

$$l(\mathbf{i}) = (-1)^{i_1} + \ldots + (-1)^{i_r}, \quad m(\mathbf{i}) = \tfrac{1}{2}(r + l(\mathbf{i})), \quad n(\mathbf{i}) = \tfrac{1}{2}(r - l(\mathbf{i})).$$

Then on recalling the definition (4.16) and considering the underlying Diophantine equations, we see that

$$T_1 \ll \sum_{p \in \mathscr{P}} \sum_{\mathbf{i} \in \mathscr{I}} \oint \mathfrak{F}_0^{\boldsymbol{\sigma}(\mathbf{i})}(\boldsymbol{\alpha};\mathbf{0}) f(\boldsymbol{\alpha};X)^{s+r-m} f(-\boldsymbol{\alpha};X)^{s+r-n} \, \mathrm{d}\boldsymbol{\alpha}.$$

An application of Schwarz's inequality therefore reveals that

$$T_1 \ll \max_{p \in \mathscr{P}} \Big(\max_{\boldsymbol{\sigma} \in \Sigma_r} \oint |\mathfrak{F}_0^{\boldsymbol{\sigma}}(\boldsymbol{\alpha};\mathbf{0})^2 f(\boldsymbol{\alpha};X)^{2s}| \, \mathrm{d}\boldsymbol{\alpha} \Big)^{1/2} \Big(\oint |f(\boldsymbol{\alpha};X)|^{2s+2r} \, \mathrm{d}\boldsymbol{\alpha} \Big)^{1/2}.$$

Recalling now (3.4) and (4.19), we deduce that there exists a prime number p with $M < p \leqslant 2M$ for which

$$T_1 \ll (I_{0,0}(X))^{1/2}(J_{s+r}(X))^{1/2}.$$

By reference to (5.9), we therefore arrive at the upper bound

$$J_{s+r}(X) = T_0 + T_1 \ll X^{-1}J_{s+r}(X) + (I_{0,0}(X))^{1/2}(J_{s+r}(X))^{1/2},$$

whence

$$J_{s+r}(X) \ll 1 + I_{0,0}(X) \ll I_{0,0}(X). \tag{5.10}$$

Our final step is to split the summation in the definition (4.15) of $\mathfrak{f}_0(\boldsymbol{\alpha}; \mathbf{0})$ into arithmetic progressions modulo p. Thus we obtain

$$\mathfrak{f}_0(\boldsymbol{\alpha}; \mathbf{0}) = \sum_{1 \leqslant \boldsymbol{\xi} \leqslant p} \mathfrak{f}_1(\boldsymbol{\alpha}; \boldsymbol{\xi}),$$

and so it follows from Hölder's inequality that

$$|\mathfrak{f}_0(\boldsymbol{\alpha}; \mathbf{0})|^{2s} \leqslant (p^d)^{2s-1} \sum_{1 \leqslant \boldsymbol{\xi} \leqslant p} |\mathfrak{f}_1(\boldsymbol{\alpha}; \boldsymbol{\xi})|^{2s}.$$

In this way we deduce from (4.17) and (4.19) that

$$I_{0,0}(X) \ll (M^d)^{2s} \max_{1 \leqslant \boldsymbol{\xi} \leqslant p} \max_{\boldsymbol{\sigma} \in \Sigma_r} \oint |\mathfrak{F}_0^{\boldsymbol{\sigma}}(\boldsymbol{\alpha}; \mathbf{0})^2 \mathfrak{f}_1(\boldsymbol{\alpha}; \boldsymbol{\xi})^{2s}| \, \mathrm{d}\boldsymbol{\alpha} \ll M^{2sd} I_{0,1}(X).$$

The proof of the lemma is made complete by substituting this last estimate into (5.10). □

This is the point at which we fix the prime number p, once and for all, in such a way that the estimate $J_{s+r}(X) \ll M^{2sd} I_{0,1}(X)$ holds. That such a choice is possible is guaranteed by the conclusion of Lemma 5.4.

6. Auxiliary congruences

The main thrust of our argument begins with a discussion of the congruences that play a critical role in what follows. We first introduce some additional notation. When $\boldsymbol{\sigma} \in \Sigma_r$, denote by $\mathscr{B}_{a,b}^{\boldsymbol{\sigma}}(\mathbf{m}; \boldsymbol{\xi}, \boldsymbol{\eta})$ the set of solutions of the system of congruences

$$\sum_{i=1}^{r} \sigma_i F_j(\mathbf{z}_i - \boldsymbol{\eta}) \equiv m_j \pmod{p^{k_j b}} \quad (1 \leqslant j \leqslant r), \tag{6.1}$$

with

$$1 \leqslant \overline{\mathbf{z}} \leqslant p^{kb}, \quad \mathbf{z}_i \equiv \boldsymbol{\xi} \pmod{p^a} \quad (1 \leqslant i \leqslant r) \tag{6.2}$$

and

$$\Delta(\mathbf{z}_1, \dots, \mathbf{z}_r) \not\equiv 0 \pmod{p^{(K-r)a+1}}. \tag{6.3}$$

The non-singularity condition (6.3) is awkward to handle directly, and so we simplify it using the condition (6.2) by means of the following lemma.

Lemma 6.1. *One has the polynomial identity*

$$\Delta(t\mathbf{z}_1 + \boldsymbol{\xi}, \dots, t\mathbf{z}_r + \boldsymbol{\xi}) = t^{K-r}\Delta(\mathbf{z}_1, \dots, \mathbf{z}_r).$$

Proof. It suffices to establish the claimed identity in the special case $t = 1$, since the homogeneity of the polynomials $F_j(\mathbf{x})$ ensures that

$$\Delta(t\mathbf{z}_1, \dots, t\mathbf{z}_r) = \Big(\prod_{j=1}^{r} t^{k_j-1}\Big)\Delta(\mathbf{z}_1, \dots, \mathbf{z}_r) = t^{K-r}\Delta(\mathbf{z}_1, \dots, \mathbf{z}_r).$$

Consider then the situation with $t = 1$, and recall the relation (2.3). By the chain rule, we find that for $1 \leqslant l \leqslant d$ one has

$$\frac{\partial \mathbf{F}}{\partial x_l}(\mathbf{x} + \boldsymbol{\xi}) = \frac{\partial}{\partial x_l}(\mathbf{F}(\mathbf{x} + \boldsymbol{\xi})) = \frac{\partial}{\partial x_l}(C(\boldsymbol{\xi})\mathbf{F}(\mathbf{x})) = C(\boldsymbol{\xi})\partial_l \mathbf{F}(\mathbf{x}).$$

The definition (4.1) of $\Delta_r(\overline{\mathbf{x}})$ therefore delivers the relation

$$\begin{aligned}\Delta(\mathbf{x}_1 + \boldsymbol{\xi}, \dots, \mathbf{x}_r + \boldsymbol{\xi}) &= \det(\partial_{\sigma(i)} F_j(\mathbf{x}_i + \boldsymbol{\xi}))_{1\leqslant i,j\leqslant r}\\ &= \det\big(C(\boldsymbol{\xi})(\partial_{\sigma(i)} F_j(\mathbf{x}_i))_{1\leqslant i,j\leqslant r}\big)\\ &= (\det C(\boldsymbol{\xi}))\Delta(\mathbf{x}_1, \dots, \mathbf{x}_r).\end{aligned}$$

But $C(\boldsymbol{\xi})$ is lower unitriangular, so that $\det C(\boldsymbol{\xi}) = 1$. We therefore conclude that

$$\Delta(\mathbf{x}_1 + \boldsymbol{\xi}, \dots, \mathbf{x}_r + \boldsymbol{\xi}) = \Delta(\mathbf{x}_1, \dots, \mathbf{x}_r),$$

and in view of our earlier discussion, the proof of the lemma is complete. □

We also require an analogue of Hensel's lemma in order to lift solutions of congruences to progressively higher moduli. A suitable version of this lifting process is implicit in the next lemma.

Lemma 6.2. *Let $f_1, \dots, f_t$ be polynomials in $\mathbb{Z}[x_1, \dots, x_t]$ with respective degrees $\kappa_1, \dots, \kappa_t$, and write*

$$J(\mathbf{f}; \mathbf{x}) = \det\left(\frac{\partial f_j}{\partial x_i}(\mathbf{x})\right)_{1\leqslant i,j\leqslant t}.$$

When ϖ is a prime number, and l is a natural number, let $\mathscr{N}(\mathbf{f}; \varpi^l)$ denote the number of solutions of the simultaneous congruences

$$f_j(x_1, \dots, x_t) \equiv 0 \ (mod\ \varpi^l) \quad (1 \leqslant j \leqslant t),$$

with $1 \leqslant x_i \leqslant \varpi^l$ $(1 \leqslant i \leqslant t)$ and $(J(\mathbf{f}; \mathbf{x}), \varpi) = 1$. Then $\mathscr{N}(\mathbf{f}; \varpi^l) \leqslant \kappa_1 \dots \kappa_t$.

Proof. This is [19, Theorem 1]. □

We are now equipped to establish the basic estimate for the number of solutions of the congruences (6.1) comprising $\mathscr{B}^{\boldsymbol{\sigma}}_{a,b}(\mathbf{m}; \boldsymbol{\xi}, \boldsymbol{\eta})$.

Lemma 6.3. *Suppose that a and b are non-negative integers with $b > a$. Then*

$$\max_{1\leqslant \boldsymbol{\xi}\leqslant p^a} \max_{1\leqslant \boldsymbol{\eta}\leqslant p^b} \max_{\boldsymbol{\sigma}\in\Sigma_r} \operatorname{card}(\mathscr{B}^{\boldsymbol{\sigma}}_{a,b}(\mathbf{m}; \boldsymbol{\xi}, \boldsymbol{\eta})) \leqslant k_1 \dots k_r p^{(kb-a)rd - K(b-a)}.$$

Proof. Consider fixed integers a and b with $0 \leqslant a < b$, a fixed r-tuple $\boldsymbol{\sigma} \in \Sigma_r$, and fixed integral d-tuples $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ with $1 \leqslant \boldsymbol{\xi} \leqslant p^a$ and $1 \leqslant \boldsymbol{\eta} \leqslant p^b$. Denote by $\mathscr{D}_1(\mathbf{n})$ the set of solutions of the system of congruences

$$\sum_{i=1}^{r} \sigma_i \mathbf{F}(\mathbf{z}_i - \boldsymbol{\eta}) \equiv \mathbf{n} \pmod{p^{kb}}, \tag{6.4}$$

with $1 \leqslant \overline{\mathbf{z}} \leqslant p^{kb}$ and

$$\mathbf{z}_i \equiv \boldsymbol{\xi}_i \pmod{p^{a+1}} \quad \text{for some} \quad \overline{\boldsymbol{\xi}} \in \Xi_a(\boldsymbol{\xi}) \quad (1 \leqslant i \leqslant r). \tag{6.5}$$

Given a fixed integral r-tuple $\mathbf{m}$, the number of r-tuples $\mathbf{n}$ with $1 \leqslant \mathbf{n} \leqslant p^{kb}$, for which $n_j \equiv m_j \pmod{p^{k_j b}}$ $(1 \leqslant j \leqslant r)$, is equal to

$$\prod_{j=1}^{r} p^{(k-k_j)b} = (p^b)^{rk-K}.$$

Then it follows from (6.1) that

$$\begin{aligned} \operatorname{card}(\mathscr{B}_{a,b}^{\boldsymbol{\sigma}}(\mathbf{m}; \boldsymbol{\xi}, \boldsymbol{\eta})) &= \sum_{\substack{1 \leqslant n_1 \leqslant p^{kb} \\ n_1 \equiv m_1 \pmod{p^{k_1 b}}}} \cdots \sum_{\substack{1 \leqslant n_r \leqslant p^{kb} \\ n_r \equiv m_r \pmod{p^{k_r b}}}} \operatorname{card}(\mathscr{D}_1(\mathbf{n})) \\ &\leqslant (p^b)^{rk-K} \max_{1 \leqslant \mathbf{n} \leqslant p^{kb}} \operatorname{card}(\mathscr{D}_1(\mathbf{n})). \end{aligned} \tag{6.6}$$

We next rewrite each variable $\mathbf{z}_i$ in the shape $\mathbf{z}_i = p^a \mathbf{y}_i + \boldsymbol{\xi}$. Notice that the hypothesis that $\mathbf{z}_i \equiv \boldsymbol{\xi}_i \pmod{p^{a+1}}$ for some $\overline{\boldsymbol{\xi}} \in \Xi_a(\boldsymbol{\xi})$, recorded in (6.5), implies that $\boldsymbol{\xi}_i = p^a \mathbf{v}_i + \boldsymbol{\xi}$ for some integral d-tuple $\mathbf{v}_i$, and further that

$$\Delta(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_r) \not\equiv 0 \pmod{p^{(K-r)a+1}}.$$

But as a consequence of Lemma 6.1, one then has

$$p^{(K-r)a} \Delta(\mathbf{v}_1, \dots, \mathbf{v}_r) = \Delta(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_r) \not\equiv 0 \pmod{p^{(K-r)a+1}},$$

whence

$$\Delta(\mathbf{v}_1, \dots, \mathbf{v}_r) \not\equiv 0 \pmod{p}.$$

However, for $1 \leqslant i \leqslant r$ one has

$$p^a \mathbf{v}_i + \boldsymbol{\xi} = \boldsymbol{\xi}_i \equiv \mathbf{z}_i = p^a \mathbf{y}_i + \boldsymbol{\xi} \pmod{p^{a+1}},$$

so that $\mathbf{y}_i \equiv \mathbf{v}_i \pmod{p}$, and hence

$$\Delta(\mathbf{y}_1, \dots, \mathbf{y}_r) \equiv \Delta(\mathbf{v}_1, \dots, \mathbf{v}_r) \not\equiv 0 \pmod{p}.$$

With the substitution $\mathbf{z}_i = p^a \mathbf{y}_i + \boldsymbol{\xi}$ in (6.4), therefore, we deduce that the set of solutions $\mathscr{D}_1(\mathbf{n})$ is in bijective correspondence with the set of solutions of the system of congruences

$$\sum_{i=1}^{r} \sigma_i \mathbf{F}(p^a \mathbf{y}_i + \boldsymbol{\xi} - \boldsymbol{\eta}) \equiv \mathbf{n} \pmod{p^{kb}}, \tag{6.7}$$

with $1 \leqslant \overline{\mathbf{y}} \leqslant p^{kb-a}$ and $\Delta_r(\overline{\mathbf{y}}) \not\equiv 0 \pmod{p}$.

Let $\overline{\mathbf{y}} = \overline{\mathbf{w}}$ be any solution of the system (6.7), if indeed such a solution exists. Then it follows that all other solutions $\overline{\mathbf{y}}$ satisfy the system of congruences

$$\sum_{i=1}^{r} \sigma_i \mathbf{F}(p^a \mathbf{y}_i + \boldsymbol{\xi} - \boldsymbol{\eta}) \equiv \sum_{i=1}^{r} \sigma_i \mathbf{F}(p^a \mathbf{w}_i + \boldsymbol{\xi} - \boldsymbol{\eta}) \pmod{p^{kb}}. \tag{6.8}$$

The translation invariance formula (2.3) implies that

$$\mathbf{F}(\mathbf{x}) = C(\boldsymbol{\xi})^{-1}(\mathbf{F}(\mathbf{x} + \boldsymbol{\xi}) - \mathbf{c}_0(\boldsymbol{\xi})),$$

in which the matrix $C(\boldsymbol{\xi})$ has determinant 1. It follows that the system of congruences (6.8) is equivalent to the new system

$$\sum_{i=1}^{r} \sigma_i \mathbf{F}(p^a \mathbf{y}_i) \equiv \sum_{i=1}^{r} \sigma_i \mathbf{F}(p^a \mathbf{w}_i) \pmod{p^{kb}}.$$

By homogeneity, moreover, this new system is in turn equivalent to

$$\sum_{i=1}^{r} \sigma_i F_j(\mathbf{y}_i) \equiv \sum_{i=1}^{r} \sigma_i F_j(\mathbf{w}_i) \pmod{p^{kb-k_j a}} \quad (1 \leqslant j \leqslant r).$$

Next, we write $\mathscr{D}_2(\mathbf{u})$ for the set of solutions of the system of congruences

$$\sum_{i=1}^{r} \sigma_i F_j(\mathbf{y}_i) \equiv u_j \pmod{p^{kb-k_j a}} \quad (1 \leqslant j \leqslant r),$$

with $1 \leqslant \overline{\mathbf{y}} \leqslant p^{kb-a}$ and $\Delta_r(\overline{\mathbf{y}}) \not\equiv 0 \pmod{p}$. Then it follows from our discussion thus far that

$$\operatorname{card}(\mathscr{D}_1(\mathbf{n})) \leqslant \max_{1 \leqslant \mathbf{u} \leqslant p^{kb-a}} \operatorname{card}(\mathscr{D}_2(\mathbf{u})). \tag{6.9}$$

Denote by $\mathscr{D}_3(\mathbf{v})$ the set of solutions of the system of congruences

$$\sum_{i=1}^{r} \sigma_i \mathbf{F}(\mathbf{y}_i) \equiv \mathbf{v} \pmod{p^{kb-a}}, \tag{6.10}$$

with $1 \leqslant \overline{\mathbf{y}} \leqslant p^{kb-a}$ and $\Delta_r(\overline{\mathbf{y}}) \not\equiv 0 \pmod{p}$. Then we have

$$\begin{aligned} \operatorname{card}(\mathscr{D}_2(\mathbf{u})) &\leqslant \sum_{\substack{v_1 \equiv u_1 \pmod{p^{kb-k_1 a}} \\ 1 \leqslant v_1 \leqslant p^{kb-a}}} \cdots \sum_{\substack{v_r \equiv u_r \pmod{p^{kb-k_r a}} \\ 1 \leqslant v_r \leqslant p^{kb-a}}} \operatorname{card}(\mathscr{D}_3(\mathbf{v})) \\ &\leqslant (p^a)^{K-r} \max_{1 \leqslant \mathbf{v} \leqslant p^{kb-a}} \operatorname{card}(\mathscr{D}_3(\mathbf{v})). \end{aligned}$$

Combining this estimate with (6.9) and (6.6), we derive the upper bound

$$\operatorname{card}(\mathscr{B}^{\boldsymbol{\sigma}}_{a,b}(\mathbf{m}; \boldsymbol{\xi}, \boldsymbol{\eta})) \leqslant (p^b)^{rk-K}(p^a)^{K-r} \max_{1 \leqslant \mathbf{v} \leqslant p^{kb-a}} \operatorname{card}(\mathscr{D}_3(\mathbf{v})). \tag{6.11}$$

It is at this point that we prepare to apply Lemma 6.2. For $1 \leqslant i \leqslant r$, we consider a fixed choice for the $d-1$ coordinates y_{ij} with $j \neq \sigma(i)$. We then

define the polynomials

$$f_j(y_{1,\sigma(1)},\dots,y_{r,\sigma(r)})=\sum_{i=1}^{r}\sigma_iF_j(\mathbf{y}_i)-v_j\quad(1\leqslant j\leqslant r).$$

Consider the solutions $\overline{\mathbf{y}}$ of the system (6.10) lying in $\mathscr{D}_3(\mathbf{v})$. For $1\leqslant i\leqslant r$, there are at most p^{kb-a} possible choices for each coordinate y_{ij} with $j\neq\sigma(i)$. Write $\widetilde{\mathbf{y}}=(y_{1,\sigma(1)},\dots,y_{r,\sigma(r)})$. Then in the notation of the statement of Lemma 6.2, one has

$$J(\mathbf{f};\widetilde{\mathbf{y}})=\det(\partial_{\sigma(i)}F_j(\mathbf{y}_i))_{1\leqslant i,j\leqslant r}=\Delta_r(\overline{\mathbf{y}})\not\equiv 0\pmod p.$$

Thus we may apply Lemma 6.2 to show that the number of solutions

$$y_{1,\sigma(1)},\dots,y_{r,\sigma(r)}$$

of the system of congruences

$$f_j(y_{1,\sigma(1)},\dots,y_{r,\sigma(r)})\equiv 0\pmod{p^{kb-a}}\quad(1\leqslant j\leqslant r),$$

with $1\leqslant y_{i,\sigma(i)}\leqslant p^{kb-a}$ $(1\leqslant i\leqslant r)$, is at most $k_1\dots k_r$. In this way, we conclude that

$$\begin{aligned}\operatorname{card}(\mathscr{D}_3(\mathbf{v}))&\leqslant\sum_{\substack{1\leqslant y_{1j}\leqslant p^{kb-a}\\1\leqslant j\leqslant d\\j\neq\sigma(1)}}\dots\sum_{\substack{1\leqslant y_{rj}\leqslant p^{kb-a}\\1\leqslant j\leqslant d\\j\neq\sigma(r)}}k_1\dots k_r\\&=k_1\dots k_r(p^{kb-a})^{r(d-1)}.\end{aligned}$$

Substituting this estimate into (6.11), we arrive at the upper bound

$$\operatorname{card}(\mathscr{B}^{\boldsymbol{\sigma}}_{a,b}(\mathbf{m};\boldsymbol{\xi},\boldsymbol{\eta}))\leqslant k_1\dots k_r(p^b)^{rk-K}(p^a)^{K-r}(p^{kb-a})^{r(d-1)},$$

and the conclusion of the lemma follows at once. □

7. The conditioning process

Our next step involves extracting from the mean value $I^{\boldsymbol{\sigma}}_{a,b}(X;\boldsymbol{\xi},\boldsymbol{\eta})$ associated mean values conditioned so as to avoid singular solutions of an underlying system of congruences. Although motivated by the corresponding treatment for the classical Vinogradov system in [21, §5], the less digestible singularity condition of the present work demands some modification.

Lemma 7.1. *Let a and b be integers with $b>a\geqslant 0$. Then one has*

$$I_{a,b}(X)\ll K_{a,b}(X)+M^{2s(d-1)+r-1}I_{a,b+1}(X).$$

Proof. Consider fixed integral d-tuples $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ with $1\leqslant\boldsymbol{\xi}\leqslant p^a$ and $1\leqslant\boldsymbol{\eta}\leqslant p^b$, and an r-tuple $\boldsymbol{\sigma}\in\Sigma_r$. Then on considering the underlying Diophantine system, one finds from (4.17) that $I^{\boldsymbol{\sigma}}_{a,b}(X;\boldsymbol{\xi},\boldsymbol{\eta})$ counts the number of integral solutions of the system

$$\sum_{i=1}^{r}\sigma_i(\mathbf{F}(\mathbf{x}_i)-\mathbf{F}(\mathbf{y}_i))=\sum_{l=1}^{s}(\mathbf{F}(\mathbf{v}_l)-\mathbf{F}(\mathbf{v}_{s+l})),\tag{7.1}$$

with

$$1 \leqslant \overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{\mathbf{v}} \leqslant X, \quad \mathbf{v}_l \equiv \boldsymbol{\eta} \pmod{p^b} \quad (1 \leqslant l \leqslant 2s),$$

and satisfying the property that there exist

$$(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_r) \in \Xi_a(\boldsymbol{\xi}) \quad \text{and} \quad (\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_r) \in \Xi_a(\boldsymbol{\xi})$$

for which

$$\mathbf{x}_i \equiv \boldsymbol{\xi}_i \pmod{p^{a+1}} \quad \text{and} \quad \mathbf{y}_i \equiv \boldsymbol{\zeta}_i \pmod{p^{a+1}} \quad (1 \leqslant i \leqslant r).$$

Let T_1 denote the number of integral solutions $\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{\mathbf{v}}$ of the system (7.1), counted by $I^{\sigma}_{a,b}(X; \boldsymbol{\xi}, \boldsymbol{\eta})$, satisfying the condition that for all r-tuples $(l_1, \dots, l_r)$ with

$$1 \leqslant \mathbf{l} \leqslant 2s, \tag{7.2}$$

one has

$$\Delta(\mathbf{v}_{l_1}, \dots, \mathbf{v}_{l_r}) \equiv 0 \pmod{p^{(K-r)b+1}}.$$

Also, let T_2 denote the corresponding number of solutions satisfying the condition that for some r-tuple $\mathbf{l}$ satisfying (7.2), one has

$$\Delta(\mathbf{v}_{l_1}, \dots, \mathbf{v}_{l_r}) \not\equiv 0 \pmod{p^{(K-r)b+1}}. \tag{7.3}$$

Then we have

$$I^{\sigma}_{a,b}(X; \boldsymbol{\xi}, \boldsymbol{\eta}) \leqslant T_1 + T_2. \tag{7.4}$$

We consider first the solutions counted by T_1. Suppose that $\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{\mathbf{v}}$ is a solution counted by T_1. For each index l with $1 \leqslant l \leqslant 2s$, we may rewrite $\mathbf{v}_l$ in the shape $\mathbf{v}_l = p^b \mathbf{u}_l + \boldsymbol{\eta}$, for some integral d-tuple $\mathbf{u}_l$. Given an r-tuple $(l_1, \dots, l_r)$ satisfying (7.2), it follows from Lemma 6.1 that

$$\begin{aligned}(p^b)^{K-r} \Delta(\mathbf{u}_{l_1}, \dots, \mathbf{u}_{l_r}) &= \Delta(p^b \mathbf{u}_{l_1} + \boldsymbol{\eta}, \dots, p^b \mathbf{u}_{l_r} + \boldsymbol{\eta}) \\ &= \Delta(\mathbf{v}_{l_1}, \dots, \mathbf{v}_{l_r}) \equiv 0 \pmod{p^{(K-r)b+1}},\end{aligned}$$

whence

$$\Delta(\mathbf{u}_{l_1}, \dots, \mathbf{u}_{l_r}) \equiv 0 \pmod{p}.$$

Write $\mathscr{A} = \{1, 2, \dots, p\}$ and $\mathbb{F} = \mathbb{Z}/p\mathbb{Z}$. Then it follows that $\overline{\mathbf{u}} \equiv \overline{\boldsymbol{\nu}} \pmod{p}$ for some $\overline{\boldsymbol{\nu}} \in \mathscr{S}_{2s}(\mathscr{A}; \mathbb{F})$. Define $\mathscr{H}_b$ to be the set of $2s$-tuples $(\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_{2s})$ with $1 \leqslant \boldsymbol{\eta}_l \leqslant p^{b+1}$ $(1 \leqslant l \leqslant 2s)$ satisfying the property that $(\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_{2s}) = (\boldsymbol{\eta} + p^b \boldsymbol{\nu}_1, \dots, \boldsymbol{\eta} + p^b \boldsymbol{\nu}_{2s})$ for some $(\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_{2s}) \in \mathscr{S}_{2s}(\mathscr{A}; \mathbb{F})$. Then we have $(\mathbf{v}_1, \dots, \mathbf{v}_{2s}) \equiv (\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_{2s}) \pmod{p^{b+1}}$ for some $(\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_{2s}) \in \mathscr{H}_b$.

On considering the underlying Diophantine system, we deduce that

$$T_1 \ll \sum_{(\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_{2s}) \in \mathscr{H}_b} \oint |\mathfrak{F}^{\sigma}_a(\boldsymbol{\alpha}; \boldsymbol{\xi})|^2 |\mathfrak{f}_{b+1}(\boldsymbol{\alpha}; \boldsymbol{\eta}_1) \dots \mathfrak{f}_{b+1}(\boldsymbol{\alpha}; \boldsymbol{\eta}_{2s})| \, \mathrm{d}\boldsymbol{\alpha}.$$

In view of the elementary inequality

$$|z_1 \dots z_n| \leqslant |z_1|^n + \dots + |z_n|^n,$$

we find from (4.17) that

$$\begin{aligned} T_1 &\ll \sum_{(\boldsymbol{\eta}_1,\dots,\boldsymbol{\eta}_{2s})\in \mathscr{H}_b} \sum_{i=1}^{2s} \oint |\mathfrak{F}_a^{\boldsymbol{\sigma}}(\boldsymbol{\alpha};\boldsymbol{\xi})^2 \mathfrak{f}_{b+1}(\boldsymbol{\alpha};\boldsymbol{\eta}_i)^{2s}| \, \mathrm{d}\boldsymbol{\alpha} \\ &\ll \mathrm{card}(\mathscr{H}_b) \max_{1\leqslant \boldsymbol{\eta}_0 \leqslant p^{b+1}} I_{a,b+1}^{\boldsymbol{\sigma}}(X;\boldsymbol{\xi},\boldsymbol{\eta}_0). \end{aligned}$$

But as a consequence of Lemma 5.3, one has

$$\mathrm{card}(\mathscr{H}_b) = \mathrm{card}(\mathscr{S}_{2s}(\mathscr{A};\mathbb{F})) \ll p^{2s(d-1)+r-1}.$$

Thus we conclude from (4.19) that

$$T_1 \ll M^{2s(d-1)+r-1} I_{a,b+1}(X). \tag{7.5}$$

Next we turn our attention to the solutions $\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{\mathbf{v}}$ counted by T_2. We may suppose that for some r-tuple $\mathbf{l}$ satisfying (7.2), one has the congruence (7.3). We define τ_i for $1 \leqslant i \leqslant r$ by taking $\tau_i = 1$ when $1 \leqslant l_i \leqslant s$, and $\tau_i = -1$ when $s+1 \leqslant l_i \leqslant 2s$. Notice that since $\mathbf{v}_l \equiv \boldsymbol{\eta} \pmod{p^b}$ $(1 \leqslant l \leqslant 2s)$, the condition (7.3) implies that $(\mathbf{v}_{l_1},\dots,\mathbf{v}_{l_r}) \equiv (\boldsymbol{\nu}_{l_1},\dots,\boldsymbol{\nu}_{l_r}) \pmod{p^{b+1}}$ for some $(\boldsymbol{\nu}_{l_1},\dots,\boldsymbol{\nu}_{l_r}) \in \Xi_b(\boldsymbol{\eta})$. Thus, on considering the underlying Diophantine system, we obtain the upper bound

$$T_2 \ll \sum_{\boldsymbol{\tau}\in\Sigma_r} \oint |\mathfrak{F}_a^{\boldsymbol{\sigma}}(\boldsymbol{\alpha};\boldsymbol{\xi})^2 \mathfrak{F}_b^{\boldsymbol{\tau}}(\boldsymbol{\alpha};\boldsymbol{\eta}) \mathfrak{f}_b(\boldsymbol{\alpha};\boldsymbol{\eta})^{2s-r}| \, \mathrm{d}\boldsymbol{\alpha}.$$

On recalling that $s = ur$, an application of Hölder's inequality reveals from (4.17) and (4.18) that for some $\boldsymbol{\tau} \in \Sigma_r$, one has

$$\begin{aligned} T_2 &\ll \Big(\oint |\mathfrak{F}_a^{\boldsymbol{\sigma}}(\boldsymbol{\alpha};\boldsymbol{\xi})^2 \mathfrak{F}_b^{\boldsymbol{\tau}}(\boldsymbol{\alpha};\boldsymbol{\eta})^{2u}| \, \mathrm{d}\boldsymbol{\alpha} \Big)^{1/(2u)} \\ &\quad \times \Big(\oint |\mathfrak{F}_a^{\boldsymbol{\sigma}}(\boldsymbol{\alpha};\boldsymbol{\xi})^2 \mathfrak{f}_b(\boldsymbol{\alpha};\boldsymbol{\eta})^{2s}| \, \mathrm{d}\boldsymbol{\alpha} \Big)^{1-1/(2u)} \\ &= (K_{a,b}^{\boldsymbol{\sigma},\boldsymbol{\tau}}(X;\boldsymbol{\xi},\boldsymbol{\eta}))^{1/(2u)} (I_{a,b}^{\boldsymbol{\sigma}}(X;\boldsymbol{\xi},\boldsymbol{\eta}))^{1-1/(2u)}. \end{aligned}$$

We therefore conclude via (4.19) and (4.20) that

$$T_2 \ll (K_{a,b}(X))^{1/(2u)} (I_{a,b}(X))^{1-1/(2u)}. \tag{7.6}$$

On combining (7.5) and (7.6) with (7.4), we deduce that

$$I_{a,b}(X) \ll M^{2s(d-1)+r-1} I_{a,b+1}(X) + (K_{a,b}(X))^{1/(2u)} (I_{a,b}(X))^{1-1/(2u)},$$

whence

$$I_{a,b}(X) \ll K_{a,b}(X) + M^{2s(d-1)+r-1} I_{a,b+1}(X).$$

This completes the proof of the lemma. □

Repeated application of Lemma 7.1 shows that whenever a, b and H are non-negative integers with $b > a \geqslant 0$, then

$$I_{a,b}(X) \ll \sum_{h=0}^{H-1} (M^h)^{2s(d-1)+r-1} K_{a,b+h}(X) + (M^H)^{2s(d-1)+r-1} I_{a,b+H}(X). \tag{7.7}$$

We next show that $I_{a,b+H}(X)$ is negligible for H large enough.

Lemma 7.2. *Let a, b and H be non-negative integers with*

$$0 < b - a \leqslant H \leqslant \theta^{-1} - b.$$

Then one has

$$(M^H)^{2s(d-1)+r-1} I_{a,b+H}(X) \ll M^{-H/2}(X/M^b)^{2sd}(X/M^a)^{2rd-K+\eta_{s+r}}.$$

Proof. On considering the underlying system of Diophantine equations, we find from (4.17) that when $1 \leqslant \boldsymbol{\xi} \leqslant p^a$ and $1 \leqslant \boldsymbol{\eta} \leqslant p^{b+H}$, and $\boldsymbol{\sigma} \in \Sigma_r$, one has

$$I^{\boldsymbol{\sigma}}_{a,b+H}(X; \boldsymbol{\xi}, \boldsymbol{\eta}) \leqslant \oint |\mathfrak{f}_a(\boldsymbol{\alpha}; \boldsymbol{\xi})^{2r} \mathfrak{f}_{b+H}(\boldsymbol{\alpha}; \boldsymbol{\eta})^{2s}| \, \mathrm{d}\boldsymbol{\alpha}.$$

An application of Hölder's inequality in combination with Lemma 5.1 therefore yields the estimate

$$\begin{aligned} I^{\boldsymbol{\sigma}}_{a,b+H}(X; \boldsymbol{\xi}, \boldsymbol{\eta}) &\leqslant \Big(\oint |\mathfrak{f}_a(\boldsymbol{\alpha}; \boldsymbol{\xi})|^{2s+2r} \, \mathrm{d}\boldsymbol{\alpha} \Big)^{r/(s+r)} \\ &\quad \times \Big(\oint |\mathfrak{f}_{b+H}(\boldsymbol{\alpha}; \boldsymbol{\eta})|^{2s+2r} \, \mathrm{d}\boldsymbol{\alpha} \Big)^{s/(s+r)} \\ &\ll (J_{s+r}(2X/M^a))^{r/(s+r)} (J_{s+r}(2X/M^{b+H}))^{s/(s+r)}. \end{aligned}$$

Consequently, on recalling (4.21) and (4.24), we obtain the upper bound

$$\begin{aligned} I_{a,b+H}(X) &\ll \big((X/M^a)^{r/(s+r)} (X/M^{b+H})^{s/(s+r)} \big)^{2(s+r)d-K+\eta_{s+r}+\delta} \\ &\ll X^{\delta} (X/M^a)^{2rd-K+\eta_{s+r}} (X/M^b)^{2sd} \Upsilon, \end{aligned} \tag{7.8}$$

where

$$\Upsilon = (M^{b-a+H})^{Ks/(s+r)} M^{-2sdH}.$$

But when $H \geqslant b - a$, one has

$$\begin{aligned} H(2s(d-1) + r - 1) + (b - a + H)Ks/(s+r) - 2sdH& \\ \leqslant H(r - 1 - 2s) + 2HKs/(s+r).& \end{aligned}$$

On observing that $s \geqslant rk \geqslant K$, and hence $(s+r)^2 > K(rk+r) > Kr$, one finds that the expression $2Hs - 2HKs/(s+r)$ achieves its minimum value for $s \geqslant rk$ when $s = rk$. Hence we deduce that

$$\begin{aligned} H(2s(d-1) + r - 1) + (b - a + H)Ks/(s+r) - 2sdH& \\ \leqslant -H + H(r - 2rk) + 2HKrk/(rk+r)& \\ \leqslant -H + H(r - 2rk/(k+1)) \leqslant -H.& \end{aligned}$$

In this way, we see that for $k \geqslant 2$, one has

$$(M^H)^{2s(d-1)+r-1} \Upsilon \leqslant M^{-H},$$

whence

$$X^{\delta} (M^H)^{2s(d-1)+r-1} \Upsilon \leqslant M^{-H/2}.$$

The conclusion of the lemma follows on substituting this estimate into (7.8). □

Combining Lemma 7.2 with the upper bound (7.7), we conclude as follows.

Lemma 7.3. *Let a and b be integers with $0 \leqslant a < b$, and put $H = b - a$. Suppose that $b + H \leqslant \theta^{-1}$. Then there exists an integer h with $0 \leqslant h < H$ having the property that*

$$I_{a,b}(X) \ll (M^h)^{2s(d-1)+r-1} K_{a,b+h}(X) + M^{-H/2}(X/M^b)^{2sd}(X/M^a)^{2rd-K+\eta_{s+r}}.$$

The special case of Lemma 7.3 with $a = 0$ and $b = 1$ yields a refinement of Lemma 5.4 more easily utilised in what is to come.

Lemma 7.4. *One has $J_{s+r}(X) \ll M^{2sd} K_{0,1}(X)$.*

Proof. When $a = 0$ and $b = 1$, one has $b-a = 1$. Thus we deduce from Lemma 7.3 that

$$I_{0,1}(X) \ll K_{0,1}(X) + M^{-1/2}(X/M)^{2sd} X^{2rd-K+\eta_{s+r}}.$$

Since we may suppose that $M > X^{4\delta}$, it follows from Lemma 5.4 that

$$J_{s+r}(X) \ll M^{2sd} I_{0,1}(X) \ll M^{2sd} K_{0,1}(X) + X^{2(s+r)d-K+\eta_{s+r}-2\delta}.$$

But in view of (4.21) and (4.23), one has

$$J_{s+r}(X) \gg X^{2(s+r)d-K+\eta_{s+r}-\delta},$$

and thus we reach the upper bound

$$J_{s+r}(X) \ll M^{2sd} K_{0,1}(X) + X^{-\delta} J_{s+r}(X).$$

The conclusion of the lemma follows on disentangling this inequality. □

8. The efficient congruencing step

The mean value $K_{a,b}(X)$ contains the powerful congruence conditions which drive the iterative process. In this section we convert these conditions into a form suited for further iteration.

Lemma 8.1. *Suppose that a and b are integers with $0 \leqslant a < b \leqslant \theta^{-1}$. Then one has*

$$K_{a,b}(X) \ll M^{2(kb-a)rd-K(b-a)} (J_{s+r}(2X/M^b))^{1-r/s} (I_{b,kb}(X))^{r/s}.$$

Proof. Consider fixed r-tuples $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ with $1 \leqslant \boldsymbol{\xi} \leqslant p^a$ and $1 \leqslant \boldsymbol{\eta} \leqslant p^b$, and r-tuples $\boldsymbol{\sigma}, \boldsymbol{\tau} \in \Sigma_r$. Then on considering the underlying Diophantine system, one finds that $K_{a,b}^{\boldsymbol{\sigma},\boldsymbol{\tau}}(X; \boldsymbol{\xi}, \boldsymbol{\eta})$ counts the number of integral solutions of the system

$$\sum_{i=1}^{r} \sigma_i(\mathbf{F}(\mathbf{x}_i) - \mathbf{F}(\mathbf{y}_i)) = \sum_{l=1}^{u} \sum_{m=1}^{r} \tau_m(\mathbf{F}(\mathbf{v}_{lm}) - \mathbf{F}(\mathbf{w}_{lm})), \tag{8.1}$$

in which, for some r-tuples $\overline{\boldsymbol{\zeta}}, \overline{\boldsymbol{\nu}} \in \Xi_a(\boldsymbol{\xi})$, one has

$$1 \leqslant \overline{\mathbf{x}}, \overline{\mathbf{y}} \leqslant X, \quad \overline{\mathbf{x}} \equiv \overline{\boldsymbol{\zeta}} \pmod{p^{a+1}} \quad \text{and} \quad \overline{\mathbf{y}} \equiv \overline{\boldsymbol{\nu}} \pmod{p^{a+1}},$$

and for $1 \leqslant l \leqslant u$, for some $\overline{\boldsymbol{\mu}}_l, \overline{\boldsymbol{\theta}}_l \in \Xi_b(\boldsymbol{\eta})$, one has

$$1 \leqslant \overline{\mathbf{v}}_l, \overline{\mathbf{w}}_l \leqslant X, \quad \overline{\mathbf{v}}_l \equiv \overline{\boldsymbol{\mu}}_l \pmod{p^{b+1}} \quad \text{and} \quad \overline{\mathbf{w}}_l \equiv \overline{\boldsymbol{\theta}}_l \pmod{p^{b+1}}.$$

The translation invariance formula (2.3) implies that the system (8.1) is equivalent to the new system of equations

$$\sum_{i=1}^{r}\sigma_i(\mathbf{F}(\mathbf{x}_i-\boldsymbol{\eta})-\mathbf{F}(\mathbf{y}_i-\boldsymbol{\eta}))=\sum_{l=1}^{u}\sum_{m=1}^{r}\tau_m(\mathbf{F}(\mathbf{v}_{lm}-\boldsymbol{\eta})-\mathbf{F}(\mathbf{w}_{lm}-\boldsymbol{\eta})). \tag{8.2}$$

In any solution $\overline{\mathbf{x}},\overline{\mathbf{y}},\overline{\overline{\mathbf{v}}},\overline{\overline{\mathbf{w}}}$ counted by $K_{a,b}^{\boldsymbol{\sigma},\boldsymbol{\tau}}(X;\boldsymbol{\xi},\boldsymbol{\eta})$, one has $\overline{\mathbf{v}}_l\equiv\overline{\mathbf{w}}_l\equiv\boldsymbol{\eta}\pmod{p^b}$ $(1\leqslant l\leqslant u)$. We therefore deduce from (8.2) that

$$\sum_{i=1}^{r}\sigma_iF_j(\mathbf{x}_i-\boldsymbol{\eta})\equiv\sum_{i=1}^{r}\sigma_iF_j(\mathbf{y}_i-\boldsymbol{\eta})\pmod{p^{k_jb}}\quad(1\leqslant j\leqslant r). \tag{8.3}$$

We also have $\overline{\mathbf{x}}\equiv\overline{\mathbf{y}}\equiv\boldsymbol{\xi}\pmod{p^a}$,

$$\Delta_r(\overline{\mathbf{x}})\not\equiv 0\pmod{p^{(K-r)a+1}}\quad\text{and}\quad\Delta_r(\overline{\mathbf{y}})\not\equiv 0\pmod{p^{(K-r)a+1}}.$$

Recall the notation introduced prior to the statement of Lemma 6.1, and write

$$\mathfrak{G}_{a,b}^{\boldsymbol{\sigma}}(\boldsymbol{\alpha};\boldsymbol{\xi},\boldsymbol{\eta};\mathbf{m})=\sum_{\overline{\boldsymbol{\zeta}}\in\mathscr{B}_{a,b}^{\boldsymbol{\sigma}}(\mathbf{m};\boldsymbol{\xi},\boldsymbol{\eta})}\prod_{i=1}^{r}\mathfrak{f}_{kb}(\sigma_i\boldsymbol{\alpha};\boldsymbol{\zeta}_i).$$

On considering the underlying Diophantine system, we deduce from (8.1) and (8.3) that

$$K_{a,b}^{\boldsymbol{\sigma},\boldsymbol{\tau}}(X;\boldsymbol{\xi},\boldsymbol{\eta})=\sum_{m_1=1}^{p^{k_1b}}\dots\sum_{m_r=1}^{p^{k_rb}}\oint|\mathfrak{G}_{a,b}^{\boldsymbol{\sigma}}(\boldsymbol{\alpha};\boldsymbol{\xi},\boldsymbol{\eta};\mathbf{m})^2\mathfrak{F}_b^{\boldsymbol{\tau}}(\boldsymbol{\alpha};\boldsymbol{\eta})^{2u}|\,\mathrm{d}\boldsymbol{\alpha}. \tag{8.4}$$

By applying Cauchy's inequality in combination with the estimate supplied by Lemma 6.3, we have

$$\begin{aligned}|\mathfrak{G}_{a,b}^{\boldsymbol{\sigma}}(\boldsymbol{\alpha};\boldsymbol{\xi},\boldsymbol{\eta};\mathbf{m})|^2&\leqslant\operatorname{card}(\mathscr{B}_{a,b}^{\boldsymbol{\sigma}}(\mathbf{m};\boldsymbol{\xi},\boldsymbol{\eta}))\sum_{\overline{\boldsymbol{\zeta}}\in\mathscr{B}_{a,b}^{\boldsymbol{\sigma}}(\mathbf{m};\boldsymbol{\xi},\boldsymbol{\eta})}\prod_{i=1}^{r}|\mathfrak{f}_{kb}(\boldsymbol{\alpha};\boldsymbol{\zeta}_i)|^2\\&\ll M^{(kb-a)rd-K(b-a)}\sum_{\overline{\boldsymbol{\zeta}}\in\mathscr{B}_{a,b}^{\boldsymbol{\sigma}}(\mathbf{m};\boldsymbol{\xi},\boldsymbol{\eta})}\prod_{i=1}^{r}|\mathfrak{f}_{kb}(\boldsymbol{\alpha};\boldsymbol{\zeta}_i)|^2.\end{aligned}$$

Substituting this relation back into (8.4) and considering the underlying Diophantine system, we find that

$$\begin{aligned}&K_{a,b}^{\boldsymbol{\sigma},\boldsymbol{\tau}}(X;\boldsymbol{\xi},\boldsymbol{\eta})\\&\ll M^{(kb-a)rd-K(b-a)}\sum_{\substack{1\leqslant\overline{\boldsymbol{\zeta}}\leqslant p^{kb}\\ \overline{\boldsymbol{\zeta}}\equiv\boldsymbol{\xi}\pmod{p^a}}}\oint\Big(\prod_{i=1}^{r}|\mathfrak{f}_{kb}(\boldsymbol{\alpha};\boldsymbol{\zeta}_i)|^2\Big)|\mathfrak{F}_b^{\boldsymbol{\tau}}(\boldsymbol{\alpha};\boldsymbol{\eta})|^{2u}\,\mathrm{d}\boldsymbol{\alpha}.\end{aligned} \tag{8.5}$$

Observe next that by Hölder's inequality,

$$\sum_{\substack{1\leqslant \overline{\boldsymbol{\zeta}}\leqslant p^{kb}\\ \overline{\boldsymbol{\zeta}}\equiv \boldsymbol{\xi}\ (\mathrm{mod}\ p^a)}}\prod_{i=1}^{r}|\mathfrak{f}_{kb}(\boldsymbol{\alpha};\boldsymbol{\zeta}_i)|^2=\Bigg(\sum_{\substack{1\leqslant \boldsymbol{\zeta}\leqslant p^{kb}\\ \boldsymbol{\zeta}\equiv \boldsymbol{\xi}\ (\mathrm{mod}\ p^a)}}|\mathfrak{f}_{kb}(\boldsymbol{\alpha};\boldsymbol{\zeta})|^2\Bigg)^r$$
$$\leqslant (p^{kb-a})^{(r-1)d}\sum_{\substack{1\leqslant \boldsymbol{\zeta}\leqslant p^{kb}\\ \boldsymbol{\zeta}\equiv \boldsymbol{\xi}\ (\mathrm{mod}\ p^a)}}|\mathfrak{f}_{kb}(\boldsymbol{\alpha};\boldsymbol{\zeta})|^{2r}.$$

Then (8.5) delivers the upper bound

$$K_{a,b}^{\boldsymbol{\sigma},\boldsymbol{\tau}}(X;\boldsymbol{\xi},\boldsymbol{\eta})\ll M^{2(kb-a)rd-K(b-a)}\max_{1\leqslant \boldsymbol{\zeta}\leqslant p^{kb}}\oint|\mathfrak{f}_{kb}(\boldsymbol{\alpha};\boldsymbol{\zeta})^{2r}\mathfrak{F}_b^{\boldsymbol{\tau}}(\boldsymbol{\alpha};\boldsymbol{\eta})^{2u}|\,\mathrm{d}\boldsymbol{\alpha}.\tag{8.6}$$

Another application of Hölder's inequality yields the bound

$$\oint|\mathfrak{f}_{kb}(\boldsymbol{\alpha};\boldsymbol{\zeta})^{2r}\mathfrak{F}_b^{\boldsymbol{\tau}}(\boldsymbol{\alpha};\boldsymbol{\eta})^{2u}|\,\mathrm{d}\boldsymbol{\alpha}\leqslant U_1^{1-r/s}U_2^{r/s},$$

where

$$U_1=\oint|\mathfrak{F}_b^{\boldsymbol{\tau}}(\boldsymbol{\alpha};\boldsymbol{\eta})|^{2u+2}\,\mathrm{d}\boldsymbol{\alpha}$$

and

$$U_2=\oint|\mathfrak{F}_b^{\boldsymbol{\tau}}(\boldsymbol{\alpha};\boldsymbol{\eta})^2\mathfrak{f}_{kb}(\boldsymbol{\alpha};\boldsymbol{\zeta})^{2s}|\,\mathrm{d}\boldsymbol{\alpha}.$$

On considering the underlying Diophantine system, it follows from Lemma 5.1 that

$$U_1\leqslant\oint|\mathfrak{f}_b(\boldsymbol{\alpha};\boldsymbol{\eta})|^{2s+2r}\,\mathrm{d}\boldsymbol{\alpha}\ll J_{s+r}(2X/M^b),$$

whilst from (4.17) we have $U_2=I_{b,kb}^{\boldsymbol{\tau}}(X;\boldsymbol{\eta},\boldsymbol{\zeta})$. Thus we conclude that

$$\oint|\mathfrak{f}_{kb}(\boldsymbol{\alpha};\boldsymbol{\zeta})^{2r}\mathfrak{F}_b^{\boldsymbol{\tau}}(\boldsymbol{\alpha};\boldsymbol{\eta})^{2u}|\,\mathrm{d}\boldsymbol{\alpha}\ll(J_{s+r}(2X/M^b))^{1-r/s}(I_{b,kb}(X))^{r/s}.$$

On substituting this estimate into (8.6), the conclusion of the lemma follows. □

We conclude this section by extracting two simplified bounds that may be conveniently deployed in our iteration.

Lemma 8.2. *Suppose that a and b are integers with $0\leqslant a<b\leqslant\theta^{-1}$. Then*

$$[\![K_{a,b}(X)]\!]\ll X^{\eta_{s+r}+\delta}(M^{b-a})^K.$$

Proof. Consider fixed d-tuples $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ with $1\leqslant\boldsymbol{\xi}\leqslant p^a$ and $1\leqslant\boldsymbol{\eta}\leqslant p^b$, and r-tuples $\boldsymbol{\sigma},\boldsymbol{\tau}\in\Sigma_r$. Considering the underlying Diophantine system and then applying Hölder's inequality, we obtain

$$K_{a,b}^{\boldsymbol{\sigma},\boldsymbol{\tau}}(X;\boldsymbol{\xi},\boldsymbol{\eta})\leqslant\oint|\mathfrak{f}_a(\boldsymbol{\alpha};\boldsymbol{\xi})^{2r}\mathfrak{f}_b(\boldsymbol{\alpha};\boldsymbol{\eta})^{2s}|\,\mathrm{d}\boldsymbol{\alpha}$$
$$\leqslant\Big(\oint|\mathfrak{f}_a(\boldsymbol{\alpha};\boldsymbol{\xi})|^{2s+2r}\,\mathrm{d}\boldsymbol{\alpha}\Big)^{r/(s+r)}\Big(\oint|\mathfrak{f}_b(\boldsymbol{\alpha};\boldsymbol{\eta})|^{2s+2r}\,\mathrm{d}\boldsymbol{\alpha}\Big)^{s/(s+r)}.$$

Next applying Lemma 5.1, we deduce that

$$K_{a,b}(X) \ll (J_{s+r}(2X/M^a))^{r/(s+r)}(J_{s+r}(2X/M^b))^{s/(s+r)},$$

whence by (4.22) we arrive at the upper bound

$$\begin{aligned}[[K_{a,b}(X)]] &\ll \frac{X^\delta((X/M^a)^{r/(s+r)}(X/M^b)^{s/(s+r)})^{2(s+r)d-K+\eta_{s+r}}}{(X/M^b)^{2sd}(X/M^a)^{2rd-K}}\\ &\ll X^{\eta_{s+r}+\delta}(M^{b-a})^{Ks/(s+r)}.\end{aligned}$$

The conclusion of the lemma follows. □

The basic iterative relation follows by applying Lemma 7.3 in combination with Lemma 8.1.

Lemma 8.3. *Suppose that a and b are integers with $0 \leqslant a < b \leqslant \frac{1}{2}(k\theta)^{-1}$, and put $H = (k-1)b$. Then there exists an integer h, with $0 \leqslant h < H$, having the property that*

$$\begin{aligned}[[K_{a,b}(X)]] \ll{}& X^\delta M^{-(2s-r+1)hr/s}(X/M^b)^{\eta_{s+r}(1-r/s)}[[K_{b,kb+h}(X)]]^{r/s}\\ &+ M^{-rH/(3s)}(X/M^b)^{\eta_{s+r}}.\end{aligned}$$

Proof. We deduce from Lemma 8.1 via (4.22) that

$$[[K_{a,b}(X)]] \ll (M^b)^{2sd}(M^a)^{2rd-K}M^{2(kb-a)rd-K(b-a)}T_1^{1-r/s}T_2^{r/s}, \tag{8.7}$$

where

$$T_1 = \frac{J_{s+r}(2X/M^b)}{X^{2(s+r)d-K}} \quad \text{and} \quad T_2 = \frac{I_{b,kb}(X)}{X^{2(s+r)d-K}}.$$

But

$$T_1 \ll (M^{-b})^{2(s+r)d-K}(X/M^b)^{\eta_{s+r}+\delta}. \tag{8.8}$$

Also, on putting $H = (k-1)b$, we see that

$$kb + H = (2k-1)b < \theta^{-1}.$$

Then it follows from Lemma 7.3 that there exists an integer h with $0 \leqslant h < H$ having the property that

$$T_2 \ll \frac{(M^h)^{2s(d-1)+r-1}K_{b,kb+h}(X)}{X^{2(s+r)d-K}} + \frac{M^{-H/2}(X/M^b)^{\eta_{s+r}}}{(M^{kb})^{2sd}(M^b)^{2rd-K}}.$$

Thus we see that

$$T_2 \ll (M^{-kb})^{2sd}(M^{-b})^{2rd-K}\Omega, \tag{8.9}$$

where

$$\Omega = M^{-(2s-r+1)h}[[K_{b,kb+h}(X)]] + M^{-H/2}(X/M^b)^{\eta_{s+r}}.$$

On substituting (8.8) and (8.9) into (8.7), we deduce that

$$[[K_{a,b}(X)]] \ll M^{\omega(a,b)}(X/M^b)^{(1-r/s)(\eta_{s+r}+\delta)}\Omega^{r/s},$$

where

$$\begin{aligned}\omega(a,b) ={}& 2sdb + (2rd-K)a + 2(kb-a)rd - K(b-a)\\ &- (1-r/s)(2(s+r)d-K)b - (2sdkb + (2rd-K)b)r/s.\end{aligned}$$

A little effort reveals that $\omega(a,b)=0$, and thus we obtain

$$[[K_{a,b}(X)]] \ll (M^{-H/2})^{r/s}(X/M^b)^{\eta_{s+r}+\delta(1-r/s)} + X^{\delta}M^{-(2s-r+1)hr/s}(X/M^b)^{\eta_{s+r}(1-r/s)}[[K_{b,kb+h}(X)]]^{r/s}.$$

The conclusion of the lemma follows on noting that δ is assumed small enough that $(X/M^b)^{\delta(1-r/s)} \ll M^{rH/(6s)}$. □

9. The iterative process

Beginning with an application of Lemma 7.4, which bounds $J_{s+r}(X)$ in terms of $K_{0,1}(X)$, we may apply Lemma 8.3 to bound $J_{s+r}(X)$ in terms of $K_{a,b}(X)$ for an increasing sequence of parameters a and b. Our goal in this section is to manage this process, deriving useful information from the iterations. Our first step is to extract from Lemma 8.3 a conclusion transparent enough to be applied as the basic tool in each iterative step.

Lemma 9.1. *Suppose that a and b are integers with $0 \leqslant a < b \leqslant \frac{1}{2}(k\theta)^{-1}$. Suppose in addition that there exist non-negative numbers ψ, c and γ, with $c \leqslant 3(s/r)^N$, for which*

$$X^{\eta_{s+r}(1+\psi\theta)} \ll X^{c\delta}M^{-\gamma}[[K_{a,b}(X)]]. \tag{9.1}$$

Then, for some non-negative integer h with $h \leqslant (k-1)b$, one has

$$X^{\eta_{s+r}(1+\psi'\theta)} \ll X^{c'\delta}M^{-\gamma'}[[K_{a',b'}(X)]],$$

where

$$\psi' = (s/r)\psi + (s/r-1)b, \quad \gamma' = (s/r)\gamma + (2s-r+1)h,$$
$$c' = (s/r)(c+1), \quad a' = b \quad \text{and} \quad b' = kb+h.$$

Proof. We are at liberty to assume that $c \leqslant 3(s/r)^N$ and $\delta < (Ns)^{-3N}$, so we have

$$c\delta < \tfrac{1}{3}s^{-2N} < \theta/(3s),$$

and hence $X^{c\delta} < M^{1/(3s)}$. In addition, one has $M^{1/(3s)} > X^{\delta}$. We therefore deduce from Lemma 8.3 that there exists an integer h with $0 \leqslant h < (k-1)b$ having the property that

$$[[K_{a,b}(X)]] \ll M^{-r/(3s)}X^{\eta_{s+r}} + X^{\delta}(X/M^b)^{(1-r/s)\eta_{s+r}}\left(M^{-(2s-r+1)h}[[K_{b,kb+h}(X)]]\right)^{r/s}.$$

On making use of the hypothesised bound (9.1), and employing in addition the lower bound $r \geqslant 2$ to confirm that $X^{c\delta}M^{-r/(3s)} \ll X^{-\delta}$, we therefore obtain the estimate

$$X^{\eta_{s+r}(1+\psi\theta)} \ll X^{(c+1)\delta}M^{-\gamma-(2s-r+1)rh/s}(X/M^b)^{(1-r/s)\eta_{s+r}}[[K_{b,kb+h}(X)]]^{r/s} + X^{\eta_{s+r}-\delta},$$

whence

$$X^{\eta_{s+r}(r/s+(\psi+(1-r/s)b)\theta)} \ll X^{(c+1)\delta}M^{-\gamma-(2s-r+1)rh/s}[[K_{b,kb+h}(X)]]^{r/s}.$$

The desired conclusion now follows on raising left and right hand sides here to the power s/r. □

Our final task is to analyse the growth of the parameters as the iteration proceeds, since from this we are able to extract the conclusion of Theorem 2.1.

Lemma 9.2. *Put $s = rk$. Then one has $\eta_{s+r} = 0$.*

Proof. We may suppose that $\eta_{s+r} > 0$, for otherwise there is nothing to prove. We begin by defining three sequences (a_n), (b_n), (h_n) of non-negative integers for $0 \leqslant n \leqslant N$. We put $a_0 = 0$ and $b_0 = 1$. Then, when $0 \leqslant n < N$, we fix any integer h_n with $0 \leqslant h_n \leqslant (k-1)b_n$, and then define

$$a_{n+1} = b_n \quad \text{and} \quad b_{n+1} = kb_n + h_n. \tag{9.2}$$

We next define the auxiliary sequences (ψ_n), (c_n), (γ_n) of non-negative real numbers for $0 \leqslant n \leqslant N$ by putting $\psi_0 = 0$, $c_0 = 1$, $\gamma_0 = 0$. Then, for $0 \leqslant n < N$, we define

$$\psi_{n+1} = (s/r)\psi_n + (s/r - 1)b_n, \tag{9.3}$$

$$c_{n+1} = (s/r)(c_n + 1), \tag{9.4}$$

$$\gamma_{n+1} = (s/r)\gamma_n + (2s - r + 1)h_n. \tag{9.5}$$

It is apparent that γ_n is non-negative for $n \geqslant 0$, and an inductive argument shows that for $0 \leqslant n \leqslant N$, one has

$$c_n = \frac{2s - r}{s - r}(s/r)^n - \frac{s}{s - r} \leqslant \Big(2 + \frac{1}{k-1}\Big)(s/r)^n \leqslant 3(s/r)^n.$$

We claim that a choice may be made for the sequence (h_n) in such a manner that for $0 \leqslant n \leqslant N$, one has

$$b_n < \sqrt{N}(s/r)^n \tag{9.6}$$

and

$$X^{\eta_{s+r}(1+\psi_n\theta)} \ll X^{c_n\delta} M^{-\gamma_n} [[K_{a_n,b_n}(X)]]. \tag{9.7}$$

When $n = 0$, the relation (9.6) holds by virtue of the definition of b_0, and the relation (9.7) holds as a consequence of (4.22), (4.23) and Lemma 7.4, since the latter implies that

$$X^{\eta_{s+r}-\delta} < [[J_{s+r}(X)]] \ll [[K_{0,1}(X)]].$$

Before considering larger indices n, we conduct a preliminary analysis of the recurrence relations (9.2)-(9.5). Observe that when $n \geqslant 0$, one has

$$\gamma_{n+1} = (s/r)\gamma_n + (2s - r + 1)(b_{n+1} - kb_n),$$

whence

$$\gamma_{n+1} - (2s - r + 1)b_{n+1} = (s/r)(\gamma_n - (2s - r + 1)b_n).$$

Then we deduce by induction that

$$\begin{aligned} \gamma_n &= (2s - r + 1)b_n + (s/r)^n(\gamma_0 - (2s - r + 1)b_0) \\ &= (2s - r + 1)(b_n - (s/r)^n). \end{aligned} \tag{9.8}$$

Suppose next that the desired conclusions (9.6) and (9.7) have been established for the index $n < N$. Then from (9.6) and (4.14) one has $kb_n\theta < k(s/r)^{n-N-2} < \frac{1}{2}$, whence $b_n < \frac{1}{2}(k\theta)^{-1}$. By appealing to Lemma 9.1 we deduce from (9.7) that there exists a non-negative integer h, with $h \leqslant (k-1)b_n$, for which one has the upper bound

$$X^{\eta_{s+r}(1+\psi'\theta)} \ll X^{c'\delta}M^{-\gamma'}[[K_{a',b'}(X)]], \tag{9.9}$$

where

$$\begin{aligned} a' &= b_n = a_{n+1}, \quad b' = kb_n + h, \qquad (9.10)\\ \psi' &= (s/r)\psi_n + (s/r-1)b_n = \psi_{n+1},\\ c' &= (s/r)(c_n+1) = c_{n+1},\\ \gamma' &= (s/r)\gamma_n + (2s-r+1)h. \qquad (9.11) \end{aligned}$$

Suppose, if possible, that $b' \geqslant \sqrt{N}(s/r)^{n+1}$. The relations (9.8), (9.10) and (9.11) then show that

$$\begin{aligned} \gamma' &= (s/r)(2s-r+1)(b_n - (s/r)^n) + (2s-r+1)(b'-kb_n)\\ &= (2s-r+1)(b' - (s/r)^{n+1})\\ &\geqslant (1-1/\sqrt{N})(2s-r+1)b'. \end{aligned} \tag{9.12}$$

But $b' = kb_n + h \leqslant (2k-1)b_n < \theta^{-1}$, and so it follows from Lemma 8.2 that

$$[[K_{a',b'}(X)]] \ll X^{\eta_{s+r}+\delta}(M^{b'})^K.$$

On substituting this estimate together with (9.12) into (9.9), we obtain the upper bound

$$X^{\eta_{s+r}(1+\psi_{n+1}\theta)} \ll X^{\eta_{s+r}+(c_{n+1}+1)\delta}(M^{b'})^{K-(2s-r+1)(1-1/\sqrt{N})}.$$

We recall that $c_{n+1} \leqslant 3(s/r)^{n+1}$, so that $X^{(c_{n+1}+1)\delta} < M^{1/2}$. Also,

$$K - (1-1/\sqrt{N})(2s-r+1) \leqslant rk - (2rk-r+1) + 2s/\sqrt{N} < -1.$$

Thus we obtain

$$X^{\eta_{s+r}(1+\psi_{n+1}\theta)} \ll X^{\eta_{s+r}}M^{1-b'} \ll X^{\eta_{s+r}}M^{-1}.$$

Since ψ_{n+1} and θ are both positive, we are forced to conclude that $\eta_{s+r} < 0$, contradicting our opening hypothesis. The assumption that $b' \geqslant \sqrt{N}(s/r)^{n+1}$ is therefore untenable, and so we must in fact have $b' < \sqrt{N}(s/r)^{n+1}$. We now take h_n to be the integer h at hand, so that $b' = b_{n+1}$ and $\gamma' = \gamma_{n+1}$, and thus we confirm the upper bounds (9.6) and (9.7) with n replaced by $n+1$.

We now collect together our various bounds on the parameters in question. We have (9.6) and (9.7) for $0 \leqslant n \leqslant N$, and also the bounds $c_n \leqslant 3(s/r)^n$ and $\gamma_n \geqslant 0$. Also, by induction one finds that $b_n \geqslant k^n$ and

$$\psi_{n+1} = k\psi_n + (k-1)b_n \geqslant k\psi_n + (k-1)k^n,$$

whence $\psi_n \geqslant n(k-1)k^{n-1}$. Finally, one has $b_N\theta < (r/s)^2 < \frac{1}{2}$, so that $b_N < \theta^{-1}$. By applying Lemma 8.2 in combination with (9.7), we therefore obtain the estimate

$$X^{\eta_{s+r}(1+\psi_N\theta)} \ll X^{\eta_{s+r}+(c_N+1)\delta}(M^{b_N})^K \ll X^{\eta_{s+r}+rk}.$$

Making use again of the relation $\theta = N^{-1/2}(r/s)^{N+2}$ from (4.14), we conclude that

$$\eta_{s+r} \leqslant \frac{rk}{\psi_N\theta} \leqslant \frac{\sqrt{N}rk(s/r)^{N+2}}{N(k-1)k^{N-1}} < \frac{rk^4}{\sqrt{N}}.$$

We may take N to be as large as necessary in terms of k and r, and thus η_{s+r} can be made arbitrarily small. We are therefore forced to conclude that $\eta_{s+r} = 0$, and this completes the proof of the lemma. □

The conclusion of Theorem 2.1 is an immediate consequence of Lemma 9.2, for in view of (4.21) and (4.24) the latter shows that when $s \geqslant r(k+1)$, then one has

$$J_s(X;\mathbf{F}) \ll X^{2sd-K+\varepsilon}.$$

10. Estimates of Weyl type

The extraction of estimates analogous to that of Weyl is in general difficult, owing to the complexity of the multidimensional situation. Although it would be feasible, with extra space and care, to analyse directly the exponential sum $f(\boldsymbol{\alpha};X;\mathbf{F})$ encoding quite general translation-dilation invariant systems $\mathbf{F}$, we have chosen here to instead restrict attention to the Weyl sums associated with the system described in example (b) of §2. Such Weyl sums may be applied so as to bound the apparently more general sums $f(\boldsymbol{\alpha};X;\mathbf{F})$ mentioned above, at the cost of somewhat weaker estimates. Since our upper bounds for $J_s(X;\mathbf{F})$ are essentially optimal with $s = r(k+1)$, it transpires that the slight weakening of the Weyl exponent does not impede the bulk of applications.

We stress then that throughout this section, until indicated otherwise, the system $\mathbf{F}$ is understood to be defined by

$$\mathbf{F} = (z_1^{i_1}z_2^{i_2}\dots z_d^{i_d} : 1 \leqslant |\mathbf{i}| \leqslant k),$$

and the corresponding Weyl sum $f(\boldsymbol{\alpha};X;\mathbf{F})$ we abbreviate to $f(\boldsymbol{\alpha})$. We define r and K as in (2.6) and (2.7). We must also consider the system $\mathbf{F}'$ defined by

$$\mathbf{F}' = (z_1^{i_1}z_2^{i_2}\dots z_d^{i_d} : 1 \leqslant |\mathbf{i}| \leqslant k-1).$$

This system has rank

$$r' = \binom{k+d-1}{d} - 1,$$

and weight

$$K' = \frac{d}{d+1}(r'+1)(k-1).$$

The hard work involved in estimating $f(\boldsymbol{\alpha})$ has been presented by the first author in [9, §5], though here we take the opportunity to clarify one or two issues.

Theorem 10.1. *Fix an index* $\mathbf{j}$ *with* $2 \leqslant |\mathbf{j}| \leqslant k$*, and put* $\sigma = (2r'k)^{-1}$*. Let* $\boldsymbol{\alpha} \in \mathbb{R}^r$*, and suppose that* a *and* q *are integers with* $q \geqslant 1$*,* $(a,q) = 1$ *and* $|q\alpha_{\mathbf{j}} - a| \leqslant q^{-1}$*. Then one has*

$$|f(\boldsymbol{\alpha})| \ll X^{d+\varepsilon}(q^{-1} + X^{-1} + qX^{-|\mathbf{j}|})^{\sigma}.$$

Proof. The conclusion of the theorem is essentially immediate from [9, Theorem 5.1]. We apply Theorem 2.1 to show that when $s = r'k$, then one has

$$J_s(X; \mathbf{F}') \ll X^{2sd-K'+\Delta},$$

with $\Delta \leqslant \varepsilon$. Since $\sigma = 1/(2s)$, the upper bound for $|f(\boldsymbol{\alpha})|$ follows from the aforementioned estimate [9, Theorem 5.1]. □

We add a few words of clarification to the proof of the latter conclusion in order to serve our purposes in the proof of Theorem 10.2. The index $\mathbf{j} = (j_1, \dots, j_d)$ contains at least one coordinate j_l satisfying $j_l \geqslant 1$. By relabelling variables, if necessary, one may suppose that $l = 1$. At the top of page 24 of [9], it is asserted that there is no loss of generality in supposing that $j_1 \geqslant 1$, an assertion that is made otiose given our relabelling of variables. However, it is apparent that the argument of this proof may be reorganised so that relabelling is unnecessary, with the central elements of the argument focused on the index j_l instead of j_1. The only issue to stress is that the set $\mathscr{M} \subseteq [1, N] \cap \mathbb{Z}$ prepared at the end of the proof of [9, Theorem 5.1] now depends on l.

Our next estimate refines Theorem 10.1 so that many coefficients are approximated simultaneously with control over a common denominator. In this context, when $\theta \in \mathbb{R}$, we define $\|\theta\| = \min_{y \in \mathbb{Z}} |\theta - y|$.

Theorem 10.2. *Let* ν *be a positive number, let* A *be a real number satisfying* $1 \leqslant A \leqslant X^d$*, and write* $Q = X^{\nu}(X^dA^{-1})^{2r'k}$*. Suppose that* $|f(\boldsymbol{\alpha})| \geqslant A$*, and that* $Q \ll X^{1-2\delta}$ *for some* $\delta > 0$*. Let* l *be an integer with* $1 \leqslant l \leqslant d$*. Then for each index* $\mathbf{j} = (j_1, \dots, j_d)$ *with* $2 \leqslant |\mathbf{j}| \leqslant k$ *and* $j_l \geqslant 1$*, there exist* $a_{\mathbf{j}} \in \mathbb{Z}$ *and* $q_{\mathbf{j}} \in \mathbb{N}$ *with* $(a_{\mathbf{j}}, q_{\mathbf{j}}) = 1$ *and* $|q_{\mathbf{j}}\alpha_{\mathbf{j}} - a_{\mathbf{j}}| \leqslant QX^{\delta-|\mathbf{j}|}$*. Moreover, the least common multiple* $q_0^{(l)}$ *of the numbers* $q_{\mathbf{j}}$ *with* $2 \leqslant |\mathbf{j}| \leqslant k$ *and* $j_l \geqslant 1$ *satisfies* $q_0^{(l)} \ll Q(\log X)^{2r'k}$ *and*

$$\|q_0^{(l)}\alpha_{\mathbf{j}}\| \ll Q^2(\log X)^{2r'k}X^{\delta-|\mathbf{j}|} \quad (2 \leqslant |\mathbf{j}| \leqslant k,\ j_l \geqslant 1).$$

Proof. We apply the argument of the proof of [9, Theorem 5.2]. Consider an index $\mathbf{j} = (j_1, \dots, j_d)$ with $2 \leqslant |\mathbf{j}| \leqslant k$, and let l be any integer with $j_l \geqslant 1$. It follows from Dirichlet's Theorem that there exist coprime integers $q_{\mathbf{j}}$ and $a_{\mathbf{j}}$ with

$$1 \leqslant q_{\mathbf{j}} \leqslant Q^{-1}X^{|\mathbf{j}|-\delta} \quad \text{and} \quad |q_{\mathbf{j}}\alpha_{\mathbf{j}} - a_{\mathbf{j}}| \leqslant QX^{\delta-|\mathbf{j}|}.$$

Keeping in mind the discussion following the proof of Theorem 10.1, we may apply the argument of the proof of [9, Theorem 5.2] to deduce from Theorem 10.1 that $q_{\mathbf{j}} \ll Q(\log X)^{2r'k} \ll X^{1-\delta}$. Now fix an integer $x \in [1, X]$, and suppose that there is an integer $y \in [1, X]$ such that

$$\|(k!)^k\alpha_{\mathbf{j}}(x - y)\| \leqslant X^{1-|\mathbf{j}|}$$

for each index $\mathbf{j}$ with $2 \leqslant |\mathbf{j}| \leqslant k$ satisfying $j_l \geqslant 1$. Then the argument of the proof of [9, Theorem 5.2] following [9, equation (5.8)] shows that $q_{\mathbf{j}}$ divides $(k!)^k a_{\mathbf{j}}(x-y)$ for each of the latter indices $\mathbf{j}$. Write $q_0^{(l)}$ for the least common multiple of the integers $q_{\mathbf{j}}$ with $2 \leqslant |\mathbf{j}| \leqslant k$ and $j_l \geqslant 1$. Then the argument of the proof of [9, Theorem 5.2] shows that

$$q_0^{(l)} \ll Q(\log X)^{2r'k}. \tag{10.1}$$

Note here that our hypothesis $j_l \geqslant 1$ permits the relevant part of the argument of [9, Theorem 5.1] to be applied successfully, reflecting the discussion above following the proof of Theorem 10.1. In addition, for $2 \leqslant |\mathbf{j}| \leqslant k$ and $j_l \geqslant 1$, one has

$$\|q_0^{(l)}\alpha_{\mathbf{j}}\| \leqslant q_0^{(l)}\|q_{\mathbf{j}}\alpha_{\mathbf{j}}\| \ll Q(\log X)^{2r'k}(QX^{\delta-|\mathbf{j}|}) = Q^2(\log X)^{2r'k}X^{\delta-|\mathbf{j}|}.$$

This completes the proof of the theorem. □

We remark that the statement of [9, Theorem 5.2] should be modified to reflect the argument concluding the above proof, so that the conclusion of [9, Theorem 5.2] asserts only that $q_0 \ll Q^d(\log P)^{2sd}$. This issue follows through the work of [9] discussing Weyl sums. In particular, the conclusion of [9, Theorem 1.2] should be modified to impose a condition of the shape $\sigma^{-1} \geqslant \frac{4}{3}(d+1)rk\log(rk)$.

The next result is established via a Baker-style "final coefficient lemma" argument (see [3, Lemma 4.6]).

Theorem 10.3. *Let k be an integer with $k \geqslant 2$, and let τ be a real number with $\tau^{-1} > (2r'k+1)(d+1)$. Suppose that $|f(\boldsymbol{\alpha})| \geqslant A \geqslant X^{d-\tau+\nu}$ for some $\nu > 0$, and write $Y = (X^dA^{-1})^{k+\nu}$. Then there are integers $a_{\mathbf{j}}$ and q satisfying*

$$(q, \mathbf{a}) = 1, \quad 1 \leqslant q \ll Y \quad \text{and} \quad |q\alpha_{\mathbf{j}} - a_{\mathbf{j}}| \ll YX^{-|\mathbf{j}|} \quad (1 \leqslant |\mathbf{j}| \leqslant k).$$

Proof. This is immediate from the argument of the proof of [9, Theorem 5.5]. Write $W = X^{1-(d+1)\tau}$. We put $s = r'k$ and note that $\tau(2s+1)(d+1) < 1$, whence

$$(X^dA^{-1})^{2s(d+1)} \ll (X^{\tau-\nu})^{2s(d+1)} \ll X^{1-(d+1)\tau-2s(d+1)\nu} = WX^{-2s(d+1)\nu}.$$

As in the proof of [9, Theorem 5.5], we may apply Theorem 10.2 to show that for $1 \leqslant l \leqslant d$, there exist integers $q_0^{(l)}$ with the property that

$$1 \leqslant q_0^{(l)} \ll X^{\nu}(X^dA^{-1})^{2r'k}(\log X)^{2r'k} \ll W^{1/(d+1)}X^{-s\nu},$$

and satisfying the condition that whenever $2 \leqslant |\mathbf{j}| \leqslant k$ and $j_l \geqslant 1$, then

$$\|q_0^{(l)}\alpha_{\mathbf{j}}\| \ll X^{2\nu}(X^dA^{-1})^{4r'k}(\log X)^{2r'k}X^{\delta-|\mathbf{j}|} \ll W^{2/(d+1)}X^{\delta-|\mathbf{j}|}.$$

We take q_0 to be the least common multiple of $q_0^{(1)}, \dots, q_0^{(d)}$. In this way, we deduce first that

$$1 \leqslant q_0 \leqslant q_0^{(1)} \dots q_0^{(d)} \ll (W^{1/(d+1)}X^{-s\nu})^d \ll W^{d/(d+1)}.$$

Suppose next that $2 \leqslant |\mathbf{j}| \leqslant k$. Then there is some index l for which $j_l \geqslant 1$, and we have

$$\|q_0\alpha_{\mathbf{j}}\| \leqslant \Big(\prod_{\substack{1 \leqslant m \leqslant d \\ m \neq l}} q_0^{(m)} \Big) \|q_0^{(l)}\alpha_{\mathbf{j}}\| \ll (W^{1/(d+1)}X^{-s\nu})^{d-1} W^{2/(d+1)} X^{\delta - |\mathbf{j}|}.$$

In this way, we find that whenever $2 \leqslant |\mathbf{j}| \leqslant k$, then without any condition on $j_1, \dots, j_d$, one has

$$\|q_0\alpha_{\mathbf{j}}\| \ll W X^{\delta - |\mathbf{j}|}.$$

An application of [9, Lemma 5.4] completes the proof of the theorem, just as in the proof of [9, Theorem 5.5]. □

The conclusion of Theorem 1.3 follows at once from Theorem 10.3 as a special case. It seems worthwhile at this point to extract from Theorem 10.3 a conclusion that serves to estimate $f(\boldsymbol{\alpha}; X; \mathbf{F})$ for general translation-dilation invariant systems $\mathbf{F}$.

Theorem 10.4. *Let $\mathbf{F}$ be a reduced translation-dilation invariant system of polynomials having dimension d, rank r and degree k. Define the exponent μ by means of the relation*

$$\mu^{-1} = \left(2k \binom{k+d-1}{d} - 2k + 1 \right) (d+1).$$

Suppose that $|f(\boldsymbol{\alpha}; X; \mathbf{F})| \geqslant A \geqslant X^{d-\mu+\nu}$ for some $\nu > 0$. Write $Y = (X^d A^{-1})^{k+\nu}$. Then there are integers a_j and q, satisfying

$$(q, \mathbf{a}) = 1, \quad 1 \leqslant q \ll Y \quad \text{and} \quad |q\alpha_j - a_j| \ll Y X^{-k_j} \quad (1 \leqslant j \leqslant r).$$

Proof. By assumption, the polynomials $F_j(\mathbf{x})$ $(1 \leqslant j \leqslant r)$ are homogeneous, and satisfy the translation-dilation invariance relation (2.3) for a suitable lower unitriangular matrix $C(\boldsymbol{\xi})$. Write $\mathbf{F}$ for the column vector $(F_j(\mathbf{x}))_{1 \leqslant j \leqslant r}$, and $\mathbf{X}$ for the column vector $(\mathbf{x}^{\mathbf{i}})_{1 \leqslant |\mathbf{i}| \leqslant k}$, in which the entries are arranged in ascending colex order. Finally, put

$$\rho = \binom{k+d}{d} - 1.$$

The monomials $\mathbf{x}^{\mathbf{i}}$ span the space generated by $\{F_1, \dots, F_r\}$, and so one may write $\mathbf{F} = \mathfrak{A}\mathbf{X}$, with $\mathfrak{A}$ an $r \times \rho$ matrix having integer entries depending only on the coefficients of $\mathbf{F}$. The linear independence of the system $\mathbf{F}$ ensures that there exists an invertible $\rho \times \rho$ matrix $\mathfrak{B}$, having rational entries depending only on the coefficients of the system $\mathbf{F}$, having the property that $\mathfrak{A}\mathfrak{B}$ is an $r \times \rho$ block matrix of the shape $(O\ I_r)$, with O the $r \times (\rho - r)$ zero matrix, and I_r the $r \times r$ identity matrix. Observe that $\mathfrak{B}$ will have a block structure which associates to F_j monomials $\mathbf{x}^{\mathbf{i}}$ with $|\mathbf{i}| = k_j$.

Suppose that $|f(\boldsymbol{\alpha}; X; \mathbf{F})| \geqslant A \geqslant X^{d-\mu+\nu}$, for some $\nu > 0$. One has

$$f(\boldsymbol{\alpha}; X; \mathbf{F}) = \sum_{1 \leqslant \mathbf{x} \leqslant X} e\Big(\sum_{i=1}^{r} \alpha_i F_i(\mathbf{x}) \Big) = \sum_{1 \leqslant \mathbf{x} \leqslant X} e\Big(\sum_{1 \leqslant |\mathbf{j}| \leqslant k} \beta_{\mathbf{j}} \mathbf{x}^{\mathbf{j}} \Big),$$

wherein we have written $\boldsymbol{\beta} = \mathfrak{A}^T\boldsymbol{\alpha}$. Thus we have

$$\Big| \sum_{1 \leqslant \mathbf{x} \leqslant X} e\Big(\sum_{1 \leqslant |\mathbf{j}| \leqslant k} \beta_{\mathbf{j}} \mathbf{x}^{\mathbf{j}} \Big) \Big| \geqslant A \geqslant X^{d-\mu+\nu}.$$

It follows from Theorem 10.3 that there exist integers $c_{\mathbf{j}}$ and v, with $(v, \mathbf{c}) = 1$, satisfying

$$1 \leqslant v \ll Y \quad \text{and} \quad |v\beta_{\mathbf{j}} - c_{\mathbf{j}}| \ll YX^{-|\mathbf{j}|} \quad (1 \leqslant |\mathbf{j}| \leqslant k). \tag{10.2}$$

Now $\mathfrak{B}^T\boldsymbol{\beta} = (\mathfrak{A}\mathfrak{B})^T\boldsymbol{\alpha}$. Since $\mathfrak{A}\mathfrak{B} = (O\ I_r)$, we see that

$$(0, \dots, 0, \alpha_1, \dots, \alpha_r)^T = \mathfrak{B}^T\boldsymbol{\beta}, \tag{10.3}$$

so that $\alpha_1, \dots, \alpha_r$ are given by linear combinations of the real numbers $\beta_{\mathbf{j}}$ $(1 \leqslant |\mathbf{j}| \leqslant k)$, with rational coefficients depending at most on the coefficients of $\mathbf{F}$.

Let Ω_0 be the least natural number having the property that $\Omega_0\mathfrak{B}$ has integral entries, and let Ω be the largest of the absolute values of the entries of $\Omega_0\mathfrak{B}$. Then it follows from (10.2) and (10.3) that for $1 \leqslant i \leqslant r$, one has

$$\|\Omega_0 v\alpha_i\| \leqslant \sum_{|\mathbf{j}|=k_i} \Omega \|v\beta_{\mathbf{j}}\| \ll \rho\Omega Y X^{-k_i}.$$

Write $q_0 = v\Omega_0$ and $Z = \rho\Omega Y$. Then we find that $1 \leqslant q_0 \ll Z$ and there exist integers b_i $(1 \leqslant i \leqslant r)$ such that $|q_0\alpha_i - b_i| \ll ZX^{-k_i}$ $(1 \leqslant i \leqslant r)$. The conclusion of the theorem follows on putting $g = (q_0, \mathbf{b})$, writing $q = q_0/g$ and $a_i = b_i/g$ $(1 \leqslant i \leqslant r)$, and noting that $Z \ll Y$. □

11. Asymptotic formulae associated with Diophantine equations

The mean value estimate supplied by Theorem 2.1 may be routinely combined with estimates of Weyl type, explored in §10, so as to establish asymptotic formulae for the number of solutions of associated Diophantine systems. Such consequences have been examined already in the literature, and so our goal in this section is to sketch some conclusions, and outline the arguments necessary for their proofs. In particular, we deliberately avoid going into detail concerning proofs of these new results so as to avoid adding bulk to an already lengthy memoir.

We begin by establishing an asymptotic formula for a general counting problem of which Theorem 1.4 is essentially a special case. Consider then a reduced translation-dilation invariant system of polynomials $\mathbf{F}$ having dimension d, rank r, degree k and weight K. Let s be a natural number, and consider fixed non-zero integers c_{ij} for $1 \leqslant i \leqslant r$ and $1 \leqslant j \leqslant s$. Finally, let $N_s(X; \mathbf{F}; \mathbf{c})$ denote the number of integral solutions of the Diophantine system

$$\sum_{j=1}^{s} c_{ij} F_i(\mathbf{x}_j) = 0 \quad (1 \leqslant i \leqslant r), \tag{11.1}$$

with $1 \leqslant \overline{\mathbf{x}} \leqslant X$. For the time being, it is convenient to abbreviate $f(\boldsymbol{\alpha}; X; \mathbf{F})$ to $f(\boldsymbol{\alpha})$.

Theorem 11.1. *Suppose that $s \geqslant 2r(k+1)+1$. Suppose further that c_{ij} $(1 \leqslant i \leqslant r,\ 1 \leqslant j \leqslant s)$ are non-zero integers, and that the system (11.1) has both a non-singular real solution, and a non-singular p-adic solution for every prime p. Then there exist positive constants $\mathscr{D} = \mathscr{D}(s, \mathbf{F}, \mathbf{c})$ and $\nu = \nu(s, \mathbf{F}, \mathbf{c})$ such that*

$$N_s(X; \mathbf{F}; \mathbf{c}) = \mathscr{D}X^{sd-K} + O(X^{sd-K-\nu}).$$

Proof. We begin by defining a Hardy-Littlewood dissection. When $0 < \theta \leqslant 1$, let $\mathfrak{M}_\theta$ denote the union of the boxes

$$\mathfrak{M}_\theta(q, \mathbf{a}) = \{\boldsymbol{\alpha} \in [0,1)^r : |q\alpha_i - a_i| \leqslant X^{\theta - k_i}\ (1 \leqslant i \leqslant r)\},$$

with $1 \leqslant q \leqslant X^\theta$, $0 \leqslant \mathbf{a} \leqslant q$ and $(q, \mathbf{a}) = 1$. Complementing the *major arcs* $\mathfrak{M}_\theta$, we define the *minor arcs* $\mathfrak{m}_\theta = [0,1)^r \setminus \mathfrak{M}_\theta$. We claim that whenever $s \geqslant 2r(k+1)+1$, then for a positive number $\delta = \delta(s, \mathbf{F})$, one has

$$\int_{\mathfrak{m}_{1/3}} |f(\boldsymbol{\beta})|^s \, \mathrm{d}\boldsymbol{\beta} \ll X^{sd-K-\delta}. \tag{11.2}$$

Define $f_j(\boldsymbol{\alpha})$ to be $f(c_{1j}\alpha_1, \dots, c_{rj}\alpha_r)$. Then from the upper bound (11.2) it follows by means of Hölder's inequality that

$$\begin{aligned}\int_{\mathfrak{m}_{1/2}} \prod_{j=1}^s f_j(\boldsymbol{\alpha}) \, \mathrm{d}\boldsymbol{\alpha} &\leqslant \prod_{j=1}^s \Big(\int_{\mathfrak{m}_{1/2}} |f_j(\boldsymbol{\alpha})|^s \, \mathrm{d}\boldsymbol{\alpha} \Big)^{1/s} \\ &\ll \max_{1 \leqslant j \leqslant s} \int_{\mathfrak{m}_{1/3}} |f(\boldsymbol{\beta})|^s \, \mathrm{d}\boldsymbol{\beta} \ll X^{sd-K-\delta}. \end{aligned} \tag{11.3}$$

In order to confirm the estimate (11.2), observe first that

$$\int_{\mathfrak{m}_{1/3}} |f(\boldsymbol{\alpha})|^s \, \mathrm{d}\boldsymbol{\alpha} \leqslant \Big(\sup_{\boldsymbol{\alpha} \in \mathfrak{m}_{1/3}} |f(\boldsymbol{\alpha})| \Big)^{s-2r(k+1)} \oint |f(\boldsymbol{\alpha})|^{2r(k+1)} \, \mathrm{d}\boldsymbol{\alpha}. \tag{11.4}$$

By applying Theorem 2.1, one sees that

$$\oint |f(\boldsymbol{\alpha})|^{2r(k+1)} \, \mathrm{d}\boldsymbol{\alpha} \ll X^{2r(k+1)-K+\varepsilon}. \tag{11.5}$$

Next, define σ by means of the relation

$$\sigma^{-1} = 6(d+1)k \binom{k+d-1}{d}.$$

If we hypothesise that $|f(\boldsymbol{\alpha})| \geqslant X^{d-\sigma}$, then it follows from Theorem 10.4 that there exist integers a_i and q, with

$$1 \leqslant q \leqslant X^{1/4}, \quad (q, \mathbf{a}) = 1 \quad \text{and} \quad |q\alpha_i - a_i| \ll X^{-k_i + 1/4} \quad (1 \leqslant i \leqslant r).$$

Then one must have $\boldsymbol{\alpha} \in \mathfrak{M}_{1/3}$, and thus we infer that

$$\sup_{\boldsymbol{\alpha} \in \mathfrak{m}_{1/3}} |f(\boldsymbol{\alpha})| \leqslant X^{d-\sigma}.$$

The desired estimate (11.2) follows on combining (11.4) and (11.5), provided that we take $\delta < \sigma$. In view of our earlier discussion, this confirms the estimate (11.3).

In order to estimate the contribution of the major arcs $\mathfrak{M}_{1/2}$, we may follow the argument of [9, §6]. We avoid providing many details here. We write

$$S(q,\mathbf{a}) = \sum_{1\leqslant \mathbf{x}\leqslant q} e\Big(q^{-1}\sum_{1\leqslant i\leqslant r} a_iF_i(\mathbf{x})\Big)$$

and

$$v(\boldsymbol{\beta}) = \int_{[0,X]^d} e\Big(\sum_{1\leqslant i\leqslant r}\beta_iF_i(\boldsymbol{\gamma})\Big)\,\mathrm{d}\boldsymbol{\gamma},$$

and then put

$$V(\boldsymbol{\alpha};q,\mathbf{a}) = q^{-d}S(q,\mathbf{a})v(\boldsymbol{\alpha}-\mathbf{a}/q).$$

Define $V_j(\boldsymbol{\alpha})$ to be

$$V(c_{1j}\alpha_1,\dots,c_{rj}\alpha_r;q,c_{1j}a_1,\dots,c_{rj}a_r)$$

when $\boldsymbol{\alpha}\in\mathfrak{M}_{1/2}(q,\mathbf{a})\subseteq\mathfrak{M}_{1/2}$, and otherwise put $V_j(\boldsymbol{\alpha})=0$. Then by adapting the argument of [9, Lemma 5.3], one finds that

$$\sup_{\boldsymbol{\alpha}\in\mathfrak{M}_{1/2}}|f_j(\boldsymbol{\alpha})-V_j(\boldsymbol{\alpha})| \ll X^{d-1/2}. \tag{11.6}$$

Since $\mathrm{mes}(\mathfrak{M}_{1/2})\ll X^{(r+1)/2-K}$, it follows via Hölder's inequality that for some positive number ν, one has

$$\begin{aligned}\int_{\mathfrak{M}_{1/2}}\prod_{j=1}^s f_j(\boldsymbol{\alpha})\,\mathrm{d}\boldsymbol{\alpha} - \int_{\mathfrak{M}_{1/2}}\prod_{j=1}^s V_j(\boldsymbol{\alpha})\,\mathrm{d}\boldsymbol{\alpha}&\\ \ll X^{d-1/2}\max_{1\leqslant j\leqslant s}\int_{\mathfrak{M}_{1/2}}|V_j(\boldsymbol{\alpha})|^{s-1}\,\mathrm{d}\boldsymbol{\alpha} + X^{sd-K-\nu}.&\end{aligned}$$

Reversing course, one finds from (11.6) via Theorem 2.1 that whenever $s\geqslant 2r(k+1)+1$, then

$$\begin{aligned}\int_{\mathfrak{M}_{1/2}}|V_j(\boldsymbol{\alpha})|^{s-1}\,\mathrm{d}\boldsymbol{\alpha} &\ll \int_{\mathfrak{M}_{1/2}}|f_j(\boldsymbol{\alpha})|^{s-1}\,\mathrm{d}\boldsymbol{\alpha} + X^{(s-1)d-K-\nu}\\ &\leqslant \oint|f_j(\boldsymbol{\alpha})|^{s-1}\,\mathrm{d}\boldsymbol{\alpha} + X^{(s-1)d-K-\nu}\\ &\ll X^{(s-1)d-K+\varepsilon}.\end{aligned}$$

Thus we deduce that

$$\begin{aligned}N_s(X;\mathbf{F};\mathbf{c}) &= \int_{\mathfrak{M}_{1/2}}\prod_{j=1}^s f_j(\boldsymbol{\alpha})\,\mathrm{d}\boldsymbol{\alpha} + \int_{\mathfrak{m}_{1/2}}\prod_{j=1}^s f_j(\boldsymbol{\alpha})\,\mathrm{d}\boldsymbol{\alpha}\\ &= \int_{\mathfrak{M}_{1/2}}\prod_{j=1}^s V_j(\boldsymbol{\alpha})\,\mathrm{d}\boldsymbol{\alpha} + O(X^{sd-K-\nu}).\end{aligned} \tag{11.7}$$

The argument of the proof of [9, Theorem 6.2] is readily adapted to show that

$$\int_{\mathfrak{M}_{1/2}}\prod_{j=1}^s V_j(\boldsymbol{\alpha})\,\mathrm{d}\boldsymbol{\alpha} = \mathfrak{J}\mathfrak{S}X^{sd-K} + O(X^{sd-K-\nu}), \tag{11.8}$$

for some $\nu > 0$, where

$$\mathfrak{J} = \int_{\mathbb{R}^r} \int_{[0,1]^{sd}} e\Big(\sum_{i=1}^{r} \beta_i \sum_{j=1}^{s} c_{ij} F_i(\boldsymbol{\gamma}_j)\Big)\, \mathrm{d}\overline{\boldsymbol{\gamma}}\, \mathrm{d}\boldsymbol{\beta},$$

and

$$\mathfrak{S} = \sum_{q=1}^{\infty} \sum_{\substack{1 \leqslant \mathbf{a} \leqslant q \\ (q,\mathbf{a})=1}} q^{-sd} \prod_{j=1}^{s} S(q, c_{1j}a_1, \dots, c_{rj}a_r).$$

The absolute convergence of $\mathfrak{J}$ and $\mathfrak{S}$ follows via the methods of [9, §§5 and 6]. Here, the existence of non-singular real and p-adic solutions suffices to guarantee that $\mathfrak{J} > 0$ and $\mathfrak{S} > 0$ (see [8] and [9] for the necessary ideas). On substituting into (11.7) and (11.8), the desired conclusion follows. □

Note that, in order to count solutions of the system (11.1) with $|\overline{\mathbf{x}}| \leqslant X$, one may merely add together the contributions from the 2^d sectors accommodating the various constellations of signs amongst the d coordinates, and thus Theorem 1.4 is an immediate consequence of Theorem 11.1 corresponding to the special translation-dilation invariant system

$$\mathbf{F} = (\mathbf{x}^{\mathbf{i}} : 1 \leqslant |\mathbf{i}| \leqslant k).$$

The special case of Theorem 11.1 in which $s = 2r(k+1)+2$ and $c_{ij} = (-1)^j$ $(1 \leqslant j \leqslant s)$ delivers the following corollary.

Corollary 11.2. *Suppose that $t \geqslant r(k+1)+1$. Then there are positive numbers $\mathfrak{C} = \mathfrak{C}(t, \mathbf{F})$ and $\delta = \delta(t, \mathbf{F})$ such that*

$$J_t(X; \mathbf{F}) = \mathfrak{C} X^{2td-K} + O(X^{2td-K-\delta}).$$

We remark that the positivity of the product of singular integral and singular series, giving $\mathfrak{C} > 0$, is in this case a consequence of the lower bound provided by Theorem 3.1. Note that Theorem 1.5 is a special case of Corollary 11.2, corresponding to the same special choice of system as above.

We now move on to consider Theorem 1.6. Here we may follow the sketch provided at the end of [9, §6]. We write

$$g_j(\boldsymbol{\alpha}) = \sum_{|\mathbf{x}| \leqslant X} e\Big(\sum_{|\mathbf{i}|=k} c_j \alpha_{\mathbf{i}} \mathbf{x}^{\mathbf{i}}\Big).$$

Then it follows from Theorem 1.1 that whenever

$$s \geqslant \left(\binom{k+d}{d} - 1\right)(k+1),$$

then

$$\oint |g_j(\boldsymbol{\alpha})|^{2s}\, \mathrm{d}\boldsymbol{\alpha} \ll X^{2sd-L+\varepsilon}.$$

The argument concluding [9, §6] then suffices to prove Theorem 1.6.

Finally, we consider the translation-dilation invariant system $\mathbf{F}$, as in the preamble to Theorem 11.1, with applications in additive combinatorics in

mind. We therefore consider non-zero integers $c_1, \dots, c_s$ satisfying the condition $c_1 + \dots + c_s = 0$, and we investigate the solubility of the Diophantine system

$$c_1\mathbf{F}(\mathbf{x}_1) + \dots + c_s\mathbf{F}(\mathbf{x}_s) = \mathbf{0}. \tag{11.9}$$

We define *projected* and *subset-sum* solutions in a manner transparently analogous to that in the preamble to Theorem 1.7.

Theorem 11.3. *Suppose that $s \geqslant 2r(k+1)+1$ and $s > K + d^2$. Let c_i $(1 \leqslant i \leqslant s)$ be non-zero integers satisfying $c_1 + \dots + c_s = 0$. Suppose further that the system of equations (11.9) possesses non-singular real and p-adic solutions for each prime number p. Let $\mathscr{A} \subseteq \mathbb{Z}^d \cap [1, N]^d$, and suppose that the only solutions of the system (11.9) from $\mathscr{A}$ are either projected or subset-sum solutions. Then one has*

$$\mathrm{card}(\mathscr{A}) \ll N^d(\log\log N)^{-1/(s-1)}.$$

The conclusion of Theorem 1.7 is a special case of Theorem 11.3, as we now confirm. For in the special circumstances relevant to the statement of Theorem 1.7 one has $rk \geqslant K$, and when $k \geqslant 2$ one has in addition

$$2r + 1 \geqslant 2\binom{d+k}{k} - 1 \geqslant (d+2)(d+1) - 1 > d^2.$$

Thus we find that the hypothesis $s \geqslant 2r(k+1)+1$ already ensures that $s > K + d^2$, and Theorem 1.7 consequently follows as a direct corollary of Theorem 11.3.

The proof of Theorem 11.3 follows by adapting the methods of [11] to this more general situation wherein d may exceed 2. The conclusion of Theorem 11.1 may be employed to show that when $s \geqslant 2r(k+1)+1$ and $\mathscr{A}$ is of suitable linear uniformity with relative density δ, then the number $\mathscr{N}(\mathscr{A})$ of solutions of (11.9) with $\overline{\mathbf{x}} \in \mathscr{A}^s$ satisfies

$$\mathscr{N}(\mathscr{A}) \gg_{\mathbf{c}} \delta^s N^{sd-K}.$$

We claim that the number $\mathscr{N}_0(\mathscr{A})$ of projected solutions is $O(N^{s(d-1)+d^2})$, and that the number $\mathscr{N}_1(\mathscr{A})$ of subset-sum solutions is $O(N^{sd-K-1/s})$. Granted this claim, one finds that $\mathscr{N}(\mathscr{A}) > \mathscr{N}_0(\mathscr{A}) + \mathscr{N}_1(\mathscr{A})$ provided that

$$N^{s(d-1)+d^2} + N^{sd-K-1/s} = o(\delta^s N^{sd-K}),$$

and this relation is satisfied whenever $s > K + d^2$ and $\delta \gg (\log N)^{-1}$, for example. Under such circumstances, we conclude that the system (11.9) contains solutions from $\mathscr{A}$ that are neither projected nor subset-sum solutions. In particular, if the only solutions of (11.9) from $\mathscr{A}$ are either projected or subset-sum solutions, then $\mathrm{card}(\mathscr{A}) \ll N^d(\log N)^{-1}$. On the other hand, if $\mathscr{A}$ is not of suitable linear uniformity, then a concentration argument motivated by that of Roth [12] and described in the proof of [11, Lemma 5.3] may be adapted to this potentially higher dimensional setting to show that in the absence of non-trivial solutions, a suitable sub-progression can be obtained with higher relative density than that of $\mathscr{A}$. By iterating this argument, one deduces that

in circumstances wherein the only solutions of the system (11.9) from $\mathscr{A}$ are either projected or subset-sum solutions, then $\mathrm{card}(\mathscr{A}) \ll N^d(\log\log N)^{-1/(s-1)}$. We refer the reader to [11, §5] for details.

We have still to justify our earlier claim. We begin by counting the number of projected solutions $\mathscr{N}_0(\mathscr{A})$. Let $\overline{\mathbf{x}} = (\mathbf{x}_1, \dots, \mathbf{x}_s)$ be any projected solution counted by $\mathscr{N}_0(\mathscr{A})$. Then there exists a translate $\mathbf{a} \in \mathbb{R}^d$ with the property that $\mathrm{span}_{\mathbb{R}}\{\mathbf{x}_1 - \mathbf{a}, \dots, \mathbf{x}_s - \mathbf{a}\}$ is a vector space of dimension $m < d$. There is no loss of generality in taking $\mathbf{a} = \mathbf{x}_1$, and then $\mathrm{span}_{\mathbb{R}}\{\mathbf{x}_1 - \mathbf{a}, \dots, \mathbf{x}_s - \mathbf{a}\}$ has a basis of the shape $\mathscr{B} = \{\mathbf{x}_{i_1} - \mathbf{x}_1, \dots, \mathbf{x}_{i_m} - \mathbf{x}_1\}$ for some indices $1 < i_1 < \ldots < i_m \leqslant s$. Since $-N < \mathbf{x}_{i_l} - \mathbf{x}_1 < N$ for $1 \leqslant l \leqslant m$, one sees that the number of possible choices for $\mathscr{B}$ is $O((N^d)^m)$. But there are trivially $O(N^d)$ possible choices for $\mathbf{x}_1$, and so we find that the number T of possible translated vector spaces $\mathbf{x}_1 + \mathrm{span}_{\mathbb{R}}\mathscr{B}$ satisfies

$$T \ll N^d(N^d)^m \leqslant N^d(N^d)^{d-1} = N^{d^2}.$$

The integral vectors lying in $V = \mathrm{span}_{\mathbb{R}}\mathscr{B}$ form an integral lattice of dimension m, and hence volume considerations confirm that the number of vectors $\mathbf{y} \in \mathbb{Z}^d$ such that $-2N \leqslant \mathbf{y} \leqslant 2N$ and $\mathbf{y} \in V$ is at most $O(N^m)$. Thus we deduce that for each index i with $1 \leqslant i \leqslant s$, and each $m < d$, the number of possible choices for $\mathbf{x}_i - \mathbf{x}_1$ is at most $O(N^m)$. It follows that for fixed $\mathbf{x}_1$ and $\mathscr{B}$, the number of possible choices for $\mathbf{x}_1, \dots, \mathbf{x}_s$ is at most $O(N^{ms})$. Then the total number of possible projected solutions $\mathbf{x}_1, \dots, \mathbf{x}_s$ is

$$\mathscr{N}_0(\mathscr{A}) \leqslant N^{ms}T \ll N^{ms+d^2} \leqslant N^{(d-1)s+d^2}.$$

We turn next to the task of bounding $\mathscr{N}_1(\mathscr{A})$. The number of partitions $\{1, \dots, s\} = \mathscr{J}_1 \cup \ldots \cup \mathscr{J}_l$, with $l \geqslant 2$ and the sets $\mathscr{J}_v$ disjoint and non-empty, is plainly $O_s(1)$. Let $\mathscr{N}_2(N; \mathscr{J}_1, \dots, \mathscr{J}_l)$ denote the number of solutions of the system

$$\sum_{u \in \mathscr{J}_v} c_u \mathbf{F}(\mathbf{x}_u) = \mathbf{0} \quad (1 \leqslant v \leqslant l),$$

with $1 \leqslant \overline{\mathbf{x}} \leqslant N$. Then it follows that there exists a partition $\{1, \dots, s\} = \mathscr{J}_1 \cup \ldots \cup \mathscr{J}_l$, of the aforementioned type, for which

$$\mathscr{N}_1(\mathscr{A}) \ll \mathscr{N}_2(N; \mathscr{J}_1, \dots, \mathscr{J}_l).$$

Write $m_v = \mathrm{card}(\mathscr{J}_v)$ $(1 \leqslant v \leqslant l)$, and note that

$$m_1 + \ldots + m_l = s. \tag{11.10}$$

Finally, let n denote the number of the sets $\mathscr{J}_v$ with $1 \leqslant v \leqslant l$ satisfying the property that $\mathrm{card}(\mathscr{J}_v) = 1$.

Our next step is to consider the contribution of the subset-sum equations defined by sets $\mathscr{J}_v$, distinguishing two cases. Suppose first that $\mathrm{card}(\mathscr{J}_v) = 1$ and $\mathscr{J}_v = \{u\}$. Since we may suppose c_u to be non-zero, it follows from Lemma 5.2 that the number of solutions of the system of equations $c_u\mathbf{F}(\mathbf{x}_u) = \mathbf{0}$, with $1 \leqslant \mathbf{x}_u \leqslant N$, is $O(N^{d-1})$.

Suppose next that $\mathrm{card}(\mathscr{J}_v) > 1$. Abbreviating $f(\boldsymbol{\alpha}; N; \mathbf{F})$ to $f(\boldsymbol{\alpha})$, a trivial estimate yields

$$\oint |f(\boldsymbol{\alpha})|^2 \,\mathrm{d}\boldsymbol{\alpha} \ll N^{2d-1}.$$

On the other hand, a trivial estimate in combination with Theorem 2.1 delivers the bound

$$\oint |f(\boldsymbol{\alpha})|^s \,\mathrm{d}\boldsymbol{\alpha} \ll N^{sd-K+\varepsilon}.$$

Then it follows from Hölder's inequality and a change of variable that

$$\begin{aligned}\oint |f(c_u\boldsymbol{\alpha})|^{m_v} \,\mathrm{d}\boldsymbol{\alpha} &\leqslant \Big(\oint |f(\boldsymbol{\alpha})|^s \,\mathrm{d}\boldsymbol{\alpha}\Big)^{(m_v-2)/(s-2)} \Big(\oint |f(\boldsymbol{\alpha})|^2 \,\mathrm{d}\boldsymbol{\alpha}\Big)^{(s-m_v)/(s-2)} \\ &\ll (N^{sd-K+\varepsilon})^{(m_v-2)/(s-2)} (N^{2d-1})^{(s-m_v)/(s-2)} \\ &\ll N^{m_v d - m_v K/(s-n) - \nu_v + \varepsilon},\end{aligned}$$

where

$$\nu_v = \left(\frac{m_v-2}{s-2}\right) K - \left(\frac{m_v}{s-n}\right) K + \frac{s-m_v}{s-2}.$$

From here, a consideration of the underlying Diophantine system, followed by an application of Hölder's inequality, reveals that

$$\begin{aligned}\mathscr{N}_2(N; \mathscr{J}_1, \dots, \mathscr{J}_l) &\ll (N^{d-1})^n \prod_{\substack{1 \leqslant v \leqslant l \\ \mathrm{card}(\mathscr{J}_v)>1}} \oint \prod_{u \in \mathscr{J}_v} |f(c_u\boldsymbol{\alpha})| \,\mathrm{d}\boldsymbol{\alpha} \\ &\ll N^{(d-1)n} \prod_{\substack{1 \leqslant v \leqslant l \\ \mathrm{card}(\mathscr{J}_v)>1}} \prod_{u \in \mathscr{J}_v} \Big(\oint |f(c_u\boldsymbol{\alpha})|^{m_v} \,\mathrm{d}\boldsymbol{\alpha}\Big)^{1/m_v} \\ &\ll N^{(d-1)n} \prod_{\substack{1 \leqslant v \leqslant l \\ \mathrm{card}(\mathscr{J}_v)>1}} N^{m_v d - m_v K/(s-n) - \nu_v + \varepsilon}.\end{aligned}$$

We therefore deduce from (11.10) that

$$\mathscr{N}_1(\mathscr{A}) \ll N^{sd-K-\nu+\varepsilon},$$

where

$$\nu = n + \sum_{\substack{1 \leqslant v \leqslant l \\ \mathrm{card}(\mathscr{J}_v)>1}} \left(\left(\frac{m_v-2}{s-2}\right) K - \left(\frac{m_v}{s-n}\right) K + \frac{s-m_v}{s-2}\right).$$

On recalling (11.10), we see that

$$\nu = n + \left(\frac{(s-n)-2(l-n)}{s-2}\right) K - K + \frac{s(l-n)-(s-n)}{s-2}.$$

A modicum of computation reveals that

$$\begin{aligned}(s-2)\nu &= n(s-2) + (2-2l+n)K + s(l-n) - (s-n) \\ &= (s-2K)(l-1) + n(K-1).\end{aligned}$$

Our hypothesis on s ensures that $s \geqslant 2rk+1 \geqslant 2K+1$, and we may suppose moreover that $l \geqslant 2$. We therefore deduce that $\nu \geqslant 1/(s-2)$, and thus

$$\mathscr{N}_1(\mathscr{A}) \ll N^{sd-K-1/s}.$$

This completes the proof of our earlier claim, and hence the proof of Theorem 11.3 is now complete.

References

[1] G. I Arkhipov, V. N. Chubarikov and A. A. Karatsuba, *Trigonometric sums in number theory and analysis*, Walter de Gruyter, Berlin, 2004.

[2] G. I. Arkhipov, A. A. Karatsuba and V. N. Chubarikov, *Multiple trigonometric sums*, Trudy Mat. Inst. Steklov **151** (1980), 1–126.

[3] R. C. Baker, *Diophantine inequalities*, London Mathematical Society Monographs, vol. 1, Oxford University Press, Oxford, 1986.

[4] B. J. Birch, *Forms in many variables*, Proc. Roy. Soc. Ser. A **265** (1961/1962), 245–263.

[5] H. Davenport, *Cubic forms in sixteen variables*, Proc. Roy. Soc. Ser. A **272** (1963), 285–303.

[6] Yu. V. Linnik, *On Weyl's sums*, Mat. Sbornik (Rec. Math.) **12** (1943), 28–39.

[7] Y.-R. Liu and T. D. Wooley, *Vinogradov's mean value theorem in function fields*, in preparation.

[8] S. T. Parsell, *Multiple exponential sums over smooth numbers*, J. Reine Angew. Math. 532 (2001), 47–104.

[9] S. T. Parsell, *A generalization of Vinogradov's mean value theorem*, Proc. London Math. Soc. (3) **91** (2005), 1–32.

[10] S. T. Parsell, *Hua-type iteration for multidimensional Weyl sums*, Mathematika, in press.

[11] S. M. Prendiville, *Solution-free sets for sums of binary forms*, submitted, preprint available as arXiv:1110.1999.

[12] K. F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 104–109.

[13] W. M. Schmidt, *The density of integer points on homogeneous varieties*, Acta Math. **154** (1985), 243–296.

[14] Y. Tschinkel, *Algebraic varieties with many rational points*, Arithmetic geometry, Clay Math. Proc. **8**, Amer. Math. Soc., Providence, RI, 2009, pp. 243–334.

[15] K. Van Valckenborgh, *Squareful points of bounded height*, C. R. Math. Acad. Sci. Paris **349** (2011), 603–606.

[16] I. M. Vinogradov, *New estimates for Weyl sums*, Dokl. Akad. Nauk SSSR **8** (1935), 195–198.

[17] I. M. Vinogradov, *The method of trigonometrical sums in the theory of numbers*, Trav. Inst. Math. Steklov **23**, Moscow, 1947.

[18] T. D. Wooley, *A note on symmetric diagonal equations*, Number Theory with an emphasis on the Markoff spectrum (Provo, UT, 1991), Editors: A. D. Pollington and W. Moran, Dekker, New York, 1993, pp. 317–321.

[19] T. D. Wooley, *A note on simultaneous congruences*, J. Number Theory **58** (1996), 288–297.

[20] T. D. Wooley, *The asymptotic formula in Waring's problem*, Internat. Math. Res. Notices (2012), No. 7, 1485–1504.

[21] T. D. Wooley, *Vinogradov's mean value theorem via efficient congruencing*, Annals of Math. **175** (2012), 1575–1627.

[22] T. D. Wooley, *Vinogradov's mean value theorem via efficient congruencing, II*, submitted, preprint available as arXiv:1112.0358.

[23] T. D. Wooley, *Vinogradov's mean value theorem, and efficient congruencing in a number field*, in preparation.

STP: Department of Mathematics, West Chester University, 25 University Ave., West Chester, PA 19383, U.S.A.

E-mail address: sparsell@wcupa.edu

SMP: School of Mathematics, University of Bristol, University Walk, Clifton, Bristol BS8 1TW, United Kingdom

E-mail address: sean.prendiville@gmail.com

TDW: School of Mathematics, University of Bristol, University Walk, Clifton, Bristol BS8 1TW, United Kingdom

E-mail address: matdw@bristol.ac.uk